(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

B2

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 November 2001 (29.11.2001)

PCT

(10) International Publication Number
WO 01/90354 A1

(51) International Patent Classification[7]: C12N 15/12, 15/85, 15/86, C07K 14/47, 16/18, C12Q 1/68, G01N 33/577, A61K 31/713, 38/18, A01K 67/027

(21) International Application Number: PCT/GB01/02240

(22) International Filing Date: 21 May 2001 (21.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0012186.3 20 May 2000 (20.05.2000) GB

(71) Applicant (for all designated States except US): UNIVERSITY OF LEEDS [GB/GB]; Leeds LS2 9JT (GB).

(72) Inventors; and
(75) Inventors/Applicants (for US only): MARKHAM, Alexander, Fred [GB/GB]; Molecular Medicine Unit, Clinical Sciences Building, St. James's University Hospital, Leeds LS9 7TF (GB). JACKSON, Andrew, Peter [GB/GB]; Molecular Medicine Unit, Clinical Sciences Building, St. James's University Hospital, Leeds LS9 7TF (GB). WOODS, Christopher, Geoffrey [GB/GB]; Molecular Medicine Unit, Clinical Sciences Building, St. James's University Hospital, Leeds LS9 7TF (GB).

(74) Agent: HARRISON GODDARD FOOTE; Belgrave Hall, Belgrave Street, Leeds LS2 8DD (GB).

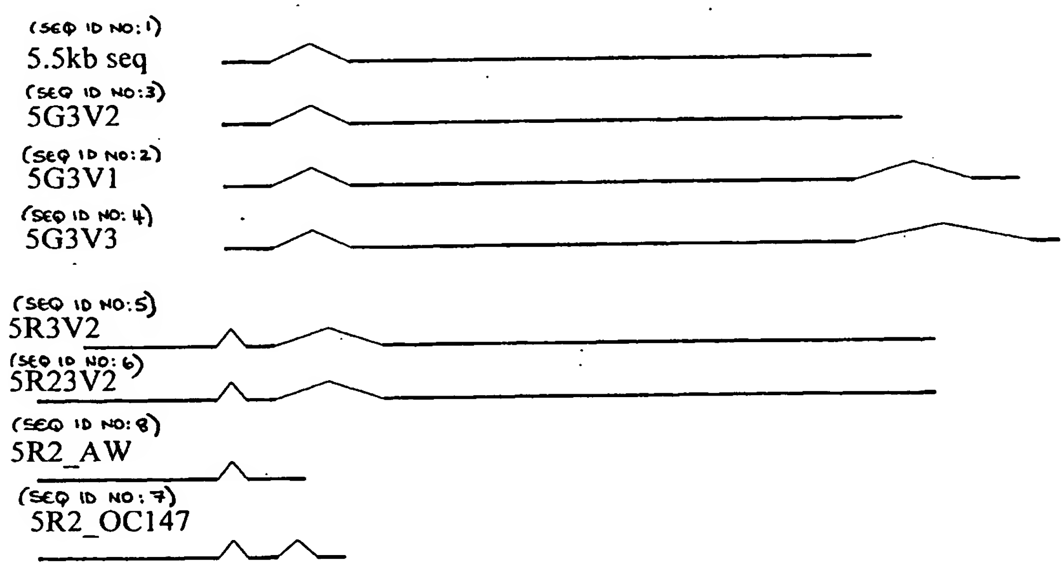
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TREATMENT OF CANCER AND NEUROLOGICAL DISEASES



(57) Abstract: The present invention relates to a nucleic acid molecule and the protein encoded thereby absence of which is associated with oral and other cancers and lack of neurogenesis. The invention also provides antibodies and the use of these products as therapeutic and/or diagnostic agents in gene therapy and/or tissue repair.

WO 01/90354 A1

Treatment of Cancer and Neurological Diseases

The present invention relates to the isolation of a nucleic acid molecule and the protein encoded thereby; antibodies raised thereto and the use of these products as therapeutic and/or diagnostic agents particularly, but not exclusively, in gene therapy and/or tissue repair such as, without limitation enhancing neuronal repair /regeneration and in the treatment of cancer.

Background to the Invention

Oral cancer has significant morbidity and mortality rates. In England and Wales the 5-year survival is around 50%. Globally, oral cancer is one of most common cancers and in some parts of the world it is the most prevalent of all cancer types. For example, in India and Sri Lanka oral cancer accounts for up to 40% of all diagnosed cancers. In addition to geographic "hot spots", there seems to be a rising trend in the increased incidence of oral cancers in many developed nations.

Recent advances in cancer management have failed to impact significantly on the outcome of oral cancer. Surgery and radiotherapy remain the principle forms of treatment with a limited role for chemotherapy. Treatment can be mutilating and is associated with high morbidity that significantly impacts on the quality of life. Speech, swallowing and taste can be markedly impaired after treatment. New treatment modalities are required for oral cancer therapy.

Statement of the Invention

We have identified a gene, from human chromosome 8p23, which is deleted in oral cancer. The gene was found to have distant similarity to the gene encoding the protein "tolloid"; and contains multiple Sushi and CUB domains. We believe that this gene may have utility in diagnosis and gene therapy applications for oral and other cancers.

Moreover, and surprisingly, the gene from human chromosome 8p23 may also be implicated in aspects of the developmental regulation of neurogenesis. We base this belief on our observations that the gene has similarity with tolloid, an important developmental gene, and the fact that it is located in the autosomal recessive microcephaly locus, MCPH1, critical region. Sequence variations in this gene can segregate with microcephaly in some families. It therefore may have utility in the diagnosis and therapy of microcephaly, as well as therapies directed to neuronal repair and regeneration, including those utilising stem cells/neural progenitor cells. Having identified this gene we believe that a further use is in the production of transgenic animals. These may have an increased predisposition to oral cancer and/or have decreased or potentially increased neocortex. Such animals would be useful not only as models of oral cancer for the evaluation of novel therapeutics but also to improve understanding of neurological developmental abnormalities. They would also serve as models to test novel therapeutics for neuronal regeneration.

According to a first aspect of the present invention there is provided an isolated nucleic acid selected from the group consisting of:

(a) DNA having the nucleotide sequence given herein as any one of SEQ ID NOS:1 TO 8;
(b) nucleic acids which hybridize to DNA of (a) above (e.g., under stringent conditions);
(c) nucleic acids having between 75-95% homology with any one of the nucleotide sequences given herein as SEQ ID NOS:1 to 8; and
(d) nucleic acids which differ from the DNA of (a), (b) or (c) above due to the degeneracy of the genetic code.

DNAs of the present invention include those coding for proteins homologous to, and having essentially the same biological properties as, the proteins disclosed herein, and particularly the DNA disclosed herein as any one of SEQ ID NOS:1 to 8 and encoding the proteins given herein as SEQ ID NOS:9 to 16 This definition is intended to encompass natural allelic variations therein. Thus, isolated DNA or

cloned genes of the present invention can be of any species of origin, including mouse, rat, rabbit, cat, porcine, and human, but are preferably of mammalian origin. Thus, DNAs which hybridize to DNA disclosed herein as any one of SEQ ID NOS:1 to 8 (or fragments or derivatives thereof which serve as hybridization probes as discussed below) and which code on expression for a protein of the present invention (e.g., a protein according to any one of SEQ ID NOS: 9 to 16), i.e. the protein lack of which is associated with oral or other cancers and/or lack of neurogenesis of the present invention are to be included in the definition.

Conditions which will permit other DNAs which code on expression for a protein of the present invention to hybridize to the DNAs of SEQ ID NO:1 to 8 disclosed herein can be determined in accordance with known techniques. For example, hybridization of such sequences may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions (e.g., conditions represented by a wash stringency of 35-40% Formamide with 5x Denhardt's solution, 0.5% SDS and 1x SSPE at 37°C; conditions represented by a wash stringency of 40-45% Formamide with 5x Denhardt's solution, 0.5% SDS, and 1x SSPE at 42°C; and conditions represented by a wash stringency of 50% Formamide with 5x Denhardt's solution, 0.5% SDS and 1x SSPE at 42°C, respectively) to DNAs of SEQ ID NO:1 to 8 disclosed herein in a standard hybridization assay. See, e.g., J. Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2d Ed. 1989) (Cold Spring Harbor Laboratory). In general, sequences which code for proteins of the present invention and which hybridize to the DNAs of SEQ ID NO:1 to 8 disclosed herein will be at least preferably 75% homologous, 85% homologous, and even 95% homologous or more with SEQ ID NO:1 to 8. Further, DNAs which code for proteins of the present invention, or DNAs which hybridize to that given as any one of SEQ ID NOS:1 to 8, but which differ in codon sequence from SEQ ID NO:1 to 8 due to the degeneracy of the genetic code, are also an aspect of this invention. The degeneracy of the genetic code, which allows different nucleic acid sequences to code for the same protein or peptide, is well known in the literature. See, e.g., U.S. Patent No. 4,757,006 to Toole et al. at Col. 2, Table 1.

According to a yet further aspect of the invention there is provided a nucleic acid molecule which encodes a protein lack of which is associated with oral or other cancers and/or lack of neurogenesis and comprises a nucleotide sequence which hybridises to the nucleic acid of any one of SEQ ID NOS:1 to 8 under high stringency conditions.

Preferably, hybridisation occurs under stringent conditions such as 1 x SSC, 0.1% SDS at 65 °C.

Preferably, the nucleic acid is mammalian in origin, for example it may be human or murine.

Preferably, the nucleic acid of the present invention is at least 2kb and up to 12 kb and may be, for example 5.5kb. The nucleic acid being located on chromosome 8p23.

According to a yet further aspect of the invention there is provided use of the nucleic acid of the present invention, in determining loss of genomic material or loss of expression of mRNA in selected target tissue(s) for diagnosing oral or other cancers and/or neurological developmental abnormalities.

According to a yet further aspect of the invention there is provided use of the nucleic acids of the present invention, in determining the presence of mutants in the DNA and thus diagnosing patients suffering from oral or other cancers and/or neurological developmental abnormalities.

According to a further aspect of the invention there is provided a polypeptide, or a protein comprising an epitope for an antibody or a protein modified by one or more amino acid modifications and comprising an epitope, or a fragment modified or unmodified comprising an eptitope for a protein lack of which is associated with oral

or other cancers and/or neurogenesis and encoded by SEQ ID NO:9 to 16. Ideally the polypeptide is encoded by the nucleic acid molecule of any one of SEQ ID NO:1 to 8.

According to a yet further aspect of the invention there is provided a polypeptide or protein encoded by the nucleic acids of the present invention, preferably the sequences of which are as set forth in SEQID NOS:9 to 16.

According to a yet further aspect of the invention there is provided a delivery vehicle comprising the isolated nucleic acid molecule or polypeptide or protein of the present invention or antibodies to these.

Reference herein to the term delivery vehicle is intended to include any vector whether a viral vector or otherwise for example, without limitation, an adenovirus, a retrovirus, a herpesvirus, a plasmid, a phage, a phagemid or a liposome.

Ideally said delivery vehicle is adapted for administration, for example, but without limitation, by suitable formulation into a suspension.

More preferably, said delivery vehicle is adapted to deliver said nucleic acid molecule or polypeptide to selected tissue. Thus the delivery vehicle is provided with means to facilitate its binding and/or penetration to a specific target site. The nature of the means comprises conventional technologies well known to those skilled in the art for example, without limitation, in the instance where the delivery vehicle is a viral vector said viral vector is provided with surface protein adapted to ensure the viral vector binds to and/or penetrates specific target tissues. Alternatively, gene expression of any one of SEQ ID NOS:1 to 8 may be under the control of a tissue specific promoter. Thus, in this way, the nucleic acid molecule or peptide, fragments or derivatives thereof of the invention can be used in gene therapy treatments.

According to a yet further aspect of the invention there is provided antibodies raised against the polypeptide, fragment or derivative thereof, of the invention. Ideally the antibodies are monoclonal and more ideally genetically engineered to be humanised. It will be apparent to those skilled in the art that the antibodies of the invention can be used to determine the expression of the polypeptide of the invention in selected target tissue and thus aid in the diagnosis of patients suffering from oral cancers and/or neurological disorders.

According to a yet further aspect of the invention there is provided use of antibodies, fragments or derivatives thereof in diagnosis/detection/identification of oral or other cancers and/or neurological disorders. It will be appreciated that the antibodies as well as the fragments or derivatives of the antibodies recognise the epitope and are capable of binding to the antigenic protein. Also useful are recombinant antibodies. The invention also includes antibodies and other compositions of matter which are specific binding partners of the polyamino acids of the present invention. Reference herein to polyamino acids is intended to include proteins and polypeptides.

The invention further provides for assays using the antibodies of the present invention to detect individuals suffering from or having a predisposition towards oral or other cancers and/or neurologiacl disorders. The assays may employ labelling, for example radioactive labels, enzymes, fluorescent compounds, chemiluminescent compounds, bioluminescent compounds and metal chelates.

Typical assays include assays known to the skilled person for quantitative or non-quantitative detection of antibodies and all involve contacting antigenic polypeptides of the present invention with a sample. The assay may involve for example and without limitation any one or more of the following techniques, RIA, EIA, ELISA, sandwich assays.

According to a yet further aspect of the invention there is provided a method for the treatment of oral cancers and/or neurological disorders comprising administering to a

patient suffering from these conditions the nucleic acid molecule or polypeptide/protein of the present invention.

Preferably, the nucleic acid molecule and/or polypeptide/protein is administered by the incorporation of said nucleic acid molecule or polypeptide/protein into a delivery vehicle as herein described and ideally the method of treatment involves the use of gene therapy.

According to a yet further aspect of the invention there is the nucleic acid and/or protein, as herein before described for use as a pharmaceutical.

According to a yet further aspect of the invention there is provided use of the nucleic acid and/or protein of the present invention for the manufacture of a medicament for the treatment of oral or other cancers and/or neurological disorders.

According to a yet further aspect of the invention there is provided a method of producing a transgenic non-human animal comprising disrupting a gene, or the effective part thereof, the gene comprising the nucleic acid of the present invention and/or the protein or effective part thereof of the present invention.

Reference herein to disruption is intended to include complete or partial disruption of expression of the protein such that the transgenic animal is unable to express levels of the said protein that are typically found in normal individuals as compared with those suffering from oral cancer and/or neurological developmental abnormalities.

Preferably, the transgenic mammal is a rodent and ideally a mouse and more preferably the gene encoding the protein lack of which is associated with oral cancer and/or neurogenesis is the nucleic acid molecule or fragment or derivative thereof as set forth in any one of SEQ ID NOS:1 to 8.

According to a yet further aspect of the invention there is provided a transgenic non-human animal whose somatic and germ cells do not contain or express a gene encoding a nucleic acid, or a nucleic acid which hybridises under high stringency conditions to, the sequence as set forth in any one of SEQ ID NOS:1 to 8, the gene having been deleted, mutated or disrupted in the animal or an ancestor of the animal at an embryonic stage and wherein the gene may be operably linked to an inducible promoter element.

Preferably, the transgenic mammal is a rodent and ideally a mouse.

According to a yet further aspect of the invention there is provided a reporter gene construct based on the promoter region of the gene, or effective part thereof, encoded by any one of SEQ ID NOS:1 to 8 i.e. the nucleic acid of the present invention.

According to a yet further aspect of the invention there is provided use of a reporter gene construct based on the promoter region of a gene, or effective part thereof, encoded by any one of SEQ ID NOS:1 to 8 in the detection/screening of pharmaceuticals and/or other compounds.

According to a yet further aspect of the invention there is provided a method of determining the presence of or predisposition towards oral or other cancers and/or neurological developmental abnormalities comprising:

(i) identifying the regions of said DNA sample that contain the nucleic acid according to the present invention;

(ii) individually hybridising parallel samples of said DNAs with oligonucleotides specific for alleles of the gene encoding any one of said nucleic acids; and

(iii) identifying from among said DNA samples those with a loss of heterozygosity for said alleles, wherein identification of a DNA sample with a loss of heterozygosity indicates presence or a predisposition towards neurological developmental abnormalities.

Preferably, the DNA sample is obtained from a human patient, alternatively RNA samples may be obtained and used in the method.

Preferably, step (i) may involve amplification of the DNA regions, typically amplification is by PCR.

Brief Description of the Figures

The invention will now be described by way of example only with reference to the following Figures wherein:

Figure 1 represents haplotypes for nine markers from 8p22-pter, for families 1 and 2 segregating autosomal recessive microcephaly. Unaffected siblings from family 1 have been omitted, for clarity. Marker order and relative distances are presented here as deduced from the Généthon map: D8S504-3cM-D8S1824-3cM-D8S1798-3cM-D8S277-2cM-D8S1819-5cM-D8S1825-13cM-D8S552-5cM-D8S1731-5cM-D8S261.

Figure 2 represents sequenced BAC's in this region from the human genome project. Position of candidate gene sequences 5R-3V2 (SEQ ID NO:5) and 5G-3V2 (SEQ ID NO:3) shown in blue (numbering corresponding to base-pair position in sequence). Sequenced BACs shown in red. BAC clone contig of [Sun, 1999 #387] shown in black, and STSs derived from this contig shown mapped onto the sequenced BACs by the vertical dashed black lines

Figure 3 represents the relationship between SEQ ID NO:1 and the sequence variants of SEQ ID NOS:2 to 8 (not to scale).

SEQ ID NO:1 to 8 represent the nucleic acids of the present invention .

SEQ ID NOS: 9 to 16 represent the corresponding protein sequences.

Materials and Methods

Subjects and Methods

A family containing five individuals affected with primary autosomal recessive microcephaly was ascertained. The family originated from the Mirpur region of Pakistan (Fig. 1, family 1). According to the clinical histories, the family confirmed that microcephaly was present from birth in all affected individuals and that there was no history of epilepsy in affected individuals. On examination, head circumferences were 5-9 SD below the population age-related mean. The affected individuals examined were 13-28 years old, and mental retardation ranged from mild to moderate in severity. None were able to read or write, but all could speak and had basic self-care skills. Except for microcephaly, there were no dysmorphic features. No affected individual had a sloping forehead, such as that described by Penrose (Cowie 1960), examination did not reveal weakness, spasticity or athertosis. Computed tomography had been performed on one affected individual at 5 years of age and results were normal. No environmental causes of microcephaly were identified. All parents appeared to be of normal intelligence and had normal head circumferences.

A further eight multiply affected consanguineous families were ascertained, with a total of 23 affected individuals displaying primary microcephaly. All of these families also originated from the Mirpur region of Pakistan and had pedigrees consistent with autosomal recessive inheritance.

DNA Extraction and Microsatellite Analysis

DNA was extracted from peripheral blood lymphocytes by means of a standard nonorganic extraction procedure. The ABI Prism linkage mapping primer set was used to perform a genomewide search. This panel contains 358 microsatellite repeat markers spaced at ~10-cM intervals, with an average heterozygosity of 0.81. PCR amplification of all the autosomal markers was performed according to the

manufacturer's specifications. Amplified markers were pooled and electrophoresed on the ABI Prism 377 gene sequencer with a 4.2% polyacrylamide gel at 3000 V and 52°C for 2 h. Fragment-length analysis was performed using the ABI Prism Genescan and Genotyper 1.1.1 analysis packages.

For fine mapping on 8p22-pter, D8S504 and D8S277 from the ABI Prism linkage set were used, and a further seven polymorphic markers from the Genome Database, were selected: tel-D8S1824-D8S1798-D8S1819-D8S1825-D8S552-D8S1731-D8S261-cen. PCR reactions were performed in 10-μl volumes that contained 50 ng genomic DNA; 1μM primers; 250μM each dGTP, dCTP, dTTP, and dATP; 5 U *Taq* DNA polymerase; and 1 x reaction buffer (1.5-2.0 mM $MgCl_2$, 10mM Tris-HCl pH 9.0, 50mM KCl, and 0.1% Triton X-100). Amplification was performed with a 5-min initial denaturing step at 95°C; 35 cycles of 94°C for 30 s, 54°C-60°C for 30 s, and 72°C for 30 s; and a final incubation step at 72°C for 5 min.

Linkage Analysis

A fully penetrant autosomal recessive mode of inheritance was assumed, and the disease allele frequency was estimated at 1/300. Two-point analysis was performed by the LINKAGE analysis programs (Terwilliger and Ott 1994) and HOMOZ-MAPMAKER was used for multipoint anlaysis (Kruglyak et al. 1995). An allele frequency of 0.1 was used in the genome screen for all markers. For further analysis of the candidate region, marker allele frequencies were calculated by genotyping 34 unrelated individuals from the same ethnic population, with a lower limit for allele frequencies set at 0.1. Heterogeneity testing was performed with the HOMOG program (Morton 1955; Terwilliger and Ott 1994).

True Microcephaly was thus mapped to chromosome 8p23 (the MCPH1 locus) (Jackson, 1998) using homozygosity mapping to perform a genomewide search. Refinement of the locus was achieved using further fluorescently labelled primers to microsatellite markers in the region. The overlap between the homozygous regions

from family 1 and 2 (Figure 1) defined the minimal critical region within which the disease gene lies, between D8S1825 and D8S1824. SEQ ID NO 1 maps to this interval on the basis of radiation hybrid mapping data (Genemap 98, Figure 4). This is additionally confirmed from genomic sequence data (SEQ ID NOS: 1 and 9) derived for the gene, which maps the gene to fully sequenced BACs (Figure 2). These BACs map to the critical region by virtue of containing polymorphic markers mapping within the critical region.

Genetic Analysis of Oral Cancers

Samples of oral cancers were obtained with local Ethics Committee approval from patients undergoing resections of their tumours. DNA was extracted from 20 such tumours and from the corresponding matched normal tissues, by standard techniques well-known in the art, providing 20 pairs of matched normal and oral cancer DNA specimens. Analysis of these paired specimens for loss of particular genetic loci in the tumours, suggestive of the local presence of a tumour suppressor gene, was performed by use of the polymerase chain reaction. Analysis of known micro-satellite markers including D8S1806, D8S1824, D8S1781, D8S1788 and D8S262 (see Figure 2) among others, showed frequent loss of one or both alleles at these loci in the majority of the oral tumours. Loss of heterozygosity was particularly frequent at the genetic markers D8S1824, D8S1781 and D8S1788.

The same matched tumour and normal tissue pairs were then compared for alterations in the gene encoding SEQ ID NO:1. In several of these tumours, deletion of both copies of this gene i.e. loss of both alleles, was detected in tumour DNA while PCR products of the expected size were amplified using DNA from matched normal control tissue. In all other cases, the relative amount of PCR amplification product generated using a variety of PCR primer pairs selected within SEQ ID NOS:1 to 8, was markedly reduced in the tumour DNA compared with that generated from normal DNA. In cases where one copy of the gene encoding the SEQ ID NO:1 was apparently retained in tumour tissue, mutations were detected in the remaining DNA

such that the open reading frame encoding the protein of SEQ ID NOS:9 to 16 was disrupted. In every case studied, the change in SEQ ID NOS:1 to 8 resulted in the alteration of a codon encoding a normal amino acid to a mis-sense amino acid or termination codon. *Thus in these cases, the oral cancer cells were unable to* synthesise the protein of SEQ ID NOS:9 to 16; as a result either of deletion of both copies of the gene described in SEQ ID NOS:1to 8 or as a result of deletion of one copy and truncating or mis-sense mutation in the residual second copy of the gene. This consistent loss of gene expression in tumours is entirely consistent with a role for the protein in SEQ ID NOS:9 to 16 as a tumour suppressor protein. It also supports the hypothesis that replacement of a functional gene by provision of the nucleic acid sequence described in SEQ ID NOS:1 to 8 would have therapeutic utility in the treatment of oral and other cancers demonstrating a similar pattern of loss of heterozygosity. Such patterns have been observed in the past for a number of other human malignancies including prostate cancer, breast cancer, ovarian cancer and colorectal cancer. Thus the nucleic acid of SEQ ID NOS:1 to 8 and/or the protein of SEQ ID NOS:9 to 16 may find equal utility in the treatment of these other common human cancers.

Accordingly the nucleic acid molecules and proteins encoded thereby of the present invention and products thereof, are of particular use in gene therapy and in identifying those suffering from or with a predisposition towards cancers, particularly oral cancers and neurological diseases.

References

1. Cowie V (1960). The genetics and sub-classification of microcephaly. J Ment. Defic. Res. 4:42-47.

2. Jackson AP, McHale DP, Campbell DA, Jafri H, Rashid Y, Mannan J, Karbani G, Corry P, Levene MI, Mueller RF, Markham AF, Lench NJ, Woods CG (1998). Primary autosomal recessive microcephaly (MCPH1) maps to chromosome 8p22-pter. Am. J. Hum. Genet. 63:541-546.

3. Morton NE (1955). The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. Am. J. Hum. Genet 7:80-96.

4. Terwilliger JD, Ott J (1994). Handbook of human genetic linkage. The Johns Hopkins University Press, Baltimore.

5. Kruglyak L, Daly MJ and Lander ES (1995). Rapid multipart linkage analysis of recessive traits in nuclear families, including homozygosity mapping. Am. J. Hum. Genet. 56:519-527.

6. Sun PC, Schmidt AP, Pashima ME, Sunwoo JB and Schlmck SB (1999). Homozygous deletions define a region of 8p23.2 containing a putative tumour suppressor gene. Genomics. 62:184-188.

P32093wo

Claims

1. An isolated nucleic acid, the nucleic acid being selected from the group consisting of:

   (a) DNAs having the nucleotide sequence given herein as any one of SEQ ID NOS:1 to 8;

   (b) nucleic acids which hybridise to DNAs of (a) above under stringent conditions;

   (c) nucleic acids having between 75-95% homology with any one of the nucleotide sequences given herein as SEQ ID NOS:1 to 8; and

   (d) nucleic acids which differ from the DNA of (a), (b) or (c) above due to the degeneracy of the genetic.

2. Nucleic acids according to claim 1 wherein the stringent conditions are 1 x SSC, 0.1% SDS at 65 °C.

3. Nucleic acids according to claim 1 consisting essentially of any one of SEQ ID NOS:1 to 8.

4. Nucleic acids according to claim 1 which hybridise to any one of SEQ ID NOS:1 to 8.

5. Nucleic acids according to claim 1 having between 75-95% homology with any one of the nucleotide sequences given herein.as SEQ ID NOS:1 to 8.

6. Nucleic acids according to claim 1 which differ from the DNAs of any one of claims 3 to 5.

7. Use of a nucleic acid according to any preceding claim in determining loss of genomic material or loss of expression of mRNA in sample.

8. Use according to claim 7 in detecting the presence of or predisposition towards oral or other cancers and/or neurological developmental abnormalities.

9. Use of a nucleic acid according to any one of claims 1 to 6 in determining the presence of mutants in DNA.

10. Use according to claim 9 in identification of patients suffering from oral or other cancers and/or neurological developmental abnormalities.

11. A polypeptide or a protein encoded by the nucleic acid molecules of any one of claims 1 to 6.

12. A delivery vehicle comprising any one of the isolated nucleic acid molecules of claims 1 to 6 or the polypeptides or proteins encoded thereby or antibodies to these polypeptides or proteins.

13. A delivery vehicle according to claim 12 comprising a viral vector selected from the group comprising an adenovirus, a retrovirus, a herpesvirus, a plasmid, a phage, a phagemid or a liposome

14. A delivery vehicle according to either claim 12 or 13 provided with surface protein adapted to facilitate binding and/or penetration to a specific target.

15. A pharmaceutical composition comprising a nucleic acid according to any one of claims 1 to 6, a polypeptide or protein according to claim 11 and/or the delivery vehicle of any one of claims 12 to 14 and a suitable excipient, diluent or carrier.

16. Antibodies which are specific binding partners of the polypeptide/protein of claim 11 or fragment or derivative thereof which are capable of binding to the antigenic part of the polypeptide/protein.

17. Antibodies according to claim 16 which are monoclonal and/or genetically engineered to be humanised.

18. Use of antibodies or antibody fragments according to either claim 16 or 17 in determining the presence or level of expression of the polypeptide or protein of claim 11.

19. Use of antibodies or antibody fragments according to either claim 16 or 17 or fragments or derivatives thereof in detecting the presence or absence of binding partners whose absence is indicative of oral or other cancers and/or neurological disorders.

20. A method for the treatment of oral cancers and/or neurological disorders comprising administering to a patient suffering from or predisposed to these conditions the nucleic acid molecule of any one of SEQ ID NOS:1 to 8 and/or the proteins encoded thereby.

21. A nucleic acid according to any one of claims 1 to 6 or polypeptide or protein of claim 11 or delivery vehicle of any one of claims 12 to 14 for use as a pharmaceutical.

22. A polyamino acid as set forth in any one of SEQ ID NOS: 9-16 for use as a pharmaceutical.

23. Use of the nucleic acids according to any one of claims 1 to 6, for the manufacture of a medicament for the treatment of oral or other cancers and/or neurological disorders.

24. A method of producing a transgenic non-human animal comprising disrupting a gene comprising the nucleic acid of any one of claims 1 to 6, or the effective part

thereof, the gene encoding a protein or effective part thereof lack of which is associated with oral or other cancers and/or lack of neurogenesis.

25. A method of producing a transgenic non-human animal comprising preventing expression of a protein or polypeptide of claim 11, or the effective part thereof, lack of expression of the protein being associated with oral or other cancers and/or lack of neurogenesis.

26. A transgenic non-human animal whose somatic and germ cells do not contain or express a gene encoding a nucleic acid according to any one of claims 1 to 6, the gene having been deleted, mutated or disrupted in the animal or an ancestor of the animal at an embryonic stage and wherein the gene may be operably linked to an inducible promoter element.

27. A transgenic non-human animal according to any one of claims 24 to 26 wherein the animal is a rodent.

28. A reporter gene construct based on the promoter region of the gene, or effective part thereof, comprising the nucleic acid of any one of claims 1 to 6.

29. Use of a reporter gene construct based on the promoter region of a gene, or effective part thereof, comprising the nucleic acid of any one of claims 1 to 6 in the detection/screening of pharmaceuticals and/or other compounds.

30. A method of determining the presence of or predisposition towards oral cancer comprising:

(i) identifying regions of a DNA sample that contain the nucleic acid according to any one of claims 1 to 6;

(ii) individually hybridising parallel samples of said DNAs with oligonucleotides specific for alleles of the gene encoding any one of said nucleic acids; and

(iii) identifying from among said DNA samples those with a loss of heterozygosity for said alleles, wherein identification of a DNA sample with a loss of heterozygosity indicates presence or a predisposition towards oral cancer.

31. A modified method according to claim 30 wherein the sample comprises RNA.

32. A method of determining the presence of or predisposition towards neurological developmental abnormalities comprising:

(i) identifying regions of a DNA sample that contain the nucleic acid according to any one of claims 1 to 6;

(ii) individually hybridising parallel samples of said DNAs with oligonucleotides specific for alleles of the gene encoding any one of said nucleic acids; and

(iii) identifying from among said DNA samples those with a loss of heterozygosity for said alleles, wherein identification of a DNA sample with a loss of heterozygosity indicates presence or a predisposition towards neurological developmental abnormalities.

33. A modified method according to claim 32 wherein the sample comprises RNA.

34. A kit comprising the nucleic acids of any one of claims 1 to 6 and a set of instructions for use thereof.

SEQ ID NO:1

cDNA sequence (partial) 5.5kb

```
ttttagggatggtatgaatttaatatttttagtattacaatatattcttataaaaaggtccaagtg
aaaaaggcgattgagttgaagtcaagaggagtcaagatgctgcccagcaaggATGGAAGCCATAAAAA
CTCTGTCTGGCATATGGAATAACATCAACCATGTGACATCCGAAGAAGATACGTTCATTATGTATCTG
GGAAAACCATGGCTTCAAGTGAAAATTCAAGTGAGCCAAGGAGGTGTTGCATTGGTCTCTGACATGTG
TCCAGATCCTGGGATTCCAGAAAATGGTAGAAGAGCAGGTTCCGACTTCAGGGTTGGTGCAAATGTAC
AGTTTTCATGTGAGGACAATTACGTGCTCCAGGGATCTAAAAGCATCACCTGTCAGAGAGTTACAGAG
ACGCTCGCTGCTTGGAGTGACCACAGGCCCATCTGCCGAGCGAGAACATGTGGATCCAATCTGCGTGG
GCCCAGCGGCGTCATTACCTCCCCTAATTATCCGGTTCAGTATGAAGATAATGCACACTGTGTGTGGG
TCATCACCACCACCGACCCGGACAAGGTCATCAAGCTTGCCTTTGAAGAGTTTGAGCTGGAGCGAGGC
TATGACACCCTGACGGTTGGTGATGCTGGGAAGGTGGGAGACACCAGATCGGTCTTGTACGTGCTCAC
GGGATCCAGTGTTCCTGACCTCATTGTGAGCATGAGCAACCAGATGTGGCTACATCTGCAGTCGGATG
ATAGCATTGGCTCACCTGGGTTTAAAGCTGTTTACCAAGAAATTGAAAAGGGAGGGTGTGGGGATCCT
GGAATCCCCGCCTATGGGAAGCGGACGGGCAGCAGTTTCCTCCATGGAGATACACTCACCTTTGAATG
CCCGGCGGCCTTTGAGCTGGTGGGGGAGAGAGTTATCACCTGTCAGCAGAACAATCAGTGGTCTGGCA
ACAAGCCCAGCTGTGTATTTTCATGTTTCTTCAACTTTACGGCATCATCTGGGATTATTCTGTCACCA
AATTATCCAGAGGAATATGGGAACAACATGAACTGTGTCTGGTTGATTATCTCGGAGCCAGGAAGTCG
AATTCACCTAATCTTTAATGATTTTGATGTTGAGCCTCAATTTGACTTTCTCGCGGTCAAGGATGATG
GCATTTCTGACATAACTGTCCTGGGTACTTTTTCTGGCAATGAAGTGCCTTCCCAGCTGGCCAGCAGT
GGGCATATAGTTCGCTTGGAATTTCAGTCTGACCATTCCACTACTGGCAGAGGGTTCAACATCACTTA
CACCACATTTGGTCAGAATGAGTGCCATGATCCTGGCATTCCTATAAACGGACGACGTTTTGGTGACA
GGTTTCTACTCGGGAGCTCGGTTTCTTTCCACTGTGATGATGGCTTTGTCAAGACCCAGGGATCCGAG
TCCATTACCTGCATACTGCAAGACGGGAACGTGGTCTGGAGCTCCACCGTGCCCCGCTGTGAAGCTCC
ATGTGGTGGACATCTGACAGCGTCCAGCGGAGTCATTTTGCCTCCTGGATGGCCAGGATATTATAAGG
ATTCTTTACATTGTGAATGGATAATTGAAGCAAAACCAGGCCACTCTATCAAAATAACTTTTGACAGA
TTTCAGACAGAGGTCAATTATGACACCTTGGAGGTCAGAGATGGGCCAGCCAGTTCGTCCCCACTGAT
CGGCGAGTACCACGGCACCCAGGCACCCCAGTTCCTCATCAGCACCGGGAACTTCATGTACCTGCTAT
TCACCACTGACAACAGCCGCTCCAGCATCGGCTTCCTCATCCACTATGAGAGTGTGACGCTTGAGTCG
GATTCCTGCCTGGACCCGGGCATCCCTGTGAACGGCCATCGCCACGGTGGAGACTTTGGCATCAGGTC
CACAGTGACTTTCAGCTGTGACCCGGGGTACACACTAAGTGACGACGAGCCCCTCGTCTGTGAGAGGA
ACCACCAGTGGAACCACGCCTTGCCCAGCTGCGACGCTCTATGTGGAGGCTACATCCAAGGGAAGAGT
GGAACAGTCCTTTCTCCTGGGTTTCCAGATTTTTATCCAAACTCTCTAAACTGCACGTGGACCATTGA
AGTGTCTCATGGGAAAGGAGTTCAAATGATCTTTCACACCTTTCATCTTGAGAGTTCCCACGACTATT
TACTGATCACAGAGGATGGAAGTTTTTCCGAGCCCGTTGCCAGGCTCACCGGGTCGGTGTTGCCTCAT
ACGATCAAGGCAGGCCTGTTTGGAAACTTCACTGCCCAGCTTCGGTTTATATCAGACTTCTCAATTTC
GTACGAGGGCTTCAATATCACATTTTCAGAATATGACCTGGAGCCATGTGATGATCCTGGAGTCCCTG
CCTTCAGCCGAAGAATTGGTTTTCACTTTGGTGTGGGAGACTCTCTGACGTTTTCCTGCTTCCTGGGA
TATCGTTTAGAAGGTGCCACCAAGCTTACCTGCCTGGGTGGGGGCCGCCGTGTGTGGAGTGCACCTCT
GCCAAGGTGTGTGGCCGAATGTGGAGCAAGTGTCAAAGGAAATGAAGGAACATTACTGTCTCCAAATT
TTCCATCCAATTATGATAATAACCATGAGTGTATCTATAAAATAGAAACAGAAGCCGGCAAGGGCATC
CACCTTAGAACACGAAGCTTCCAGCTGTTTGAAGGAGATACTCTAAAGGTATATGATGGAAAAGACAG
TTCCTCACGTCCACTGGGCACGTTCACTAAAAATGAACTTCTGGGGCTGATCCTAAACAGCACATCCA
ATCACCTGTGGCTAGAGTTCAACACCAATGGATCTGACACCGACCAAGGTTTTCAACTCACCTATACC
AGTTTTGATCTGGTAAAATGTGAGGATCCGGGCATCCCTAACTACGGCTATAGGATCCGTGATGAAGG
CCACTTTACCGACACTGTAGTTCTGTACAGTTGCAACCCGGGGTACGCCATGCATGGCAGCAACACCC
TGACCTGTTTGAGTGGAGACAGGAGAGTGTGGGACAAACCACTACCTTCGTGCATAGCGGAATGTGGT
GGTCAGATCCATGCAGCCACATCAGGACGAATATTGTCCCCTGGCTATCCAGCTCCGTATGACAACAA
CCTCCACTGCACCTGGATTATAGAGGCAGACCCAGGAAAGACCATTAGCCTCCATTTCATTGTTTTCG
ACACGGAGATGGCTCACGACATCCTCAAGGTCTGGGACGGGCCGGTGGACAGTGACATCCTGCTGAAG
GAGTGGAGTGGCTCCGCCCTTCCGGAGGACATCCACAGCACCTTCAACTCACTCACCCTGCAGTTCGA
CAGCGACTTCTTCATCAGCAAGTCTGGCTTCTCCATCCAGTTCTCCACCTCAATTGCAGCCACCTGTA
ACGATCCAGGTATGCCCCAAAATGGCACCCGCTATGGAGACAGCAGAGAGGCTGGAGACACCGTCACA
TTCCAGTGTGACCCTGGCTATCAGCTCCAAGGACAAGCCAAAATCACCTGTGTGCAGCTGAATAACCG
GTTCTTTTGGCAACCAGACCCTCCTACATGCATAGCTGCTTGTGGAGGGAATCTGACGGGCCCAGCAG
GTGTTATTTTGTCACCCAACTACCCACAGCCGTATCCTCCTGGGAAGGAATGTGACTGGAGAGTAAAA
GTGAACCCGGACTTTGTCATCGCCTTGATATTCAAAAGTTTCAACATGGAGCCCAGCTATGACTTCCT
```

ACACATCTATGAAGGGGAAGATTCCAACAGCCCCCTCATTGGGAGTTACCAGGGCTCTCAGGCCCCAG
AAAGAATAGAGAGTAGCGGAAACAGCCTGTTTCTGGCATTTCGGAGTGATGCCTCCGTGGGCCTTTCA
GGGTTCGCCATTGAATTTAAAGAGAAACCACGGGAAGCTTGTTTTGACCCAGGAAATATAATGAATGG
GACAAGAGTTGGAACAGACTTCAAGCTTGGCTCCACCATCACCTACCAGTGTGACTCTGGCTATAAGA
TTCTTGACCCCTCATCCATCACCTGTGTGATTGGGGCTGATGGGAAACCCTCCTGGGACCAAGTGCTG
CCCTCCTGCAATGCTCCCTGTGGAGGCCAGTACACGGGATCAGAAGGGGTAGTTTTATCACCAAACTA
CCCCCATAATTACACAGCTGGTCAAATATGCCTCTATTCCATCACGGTACCAAAGGAATTCGTGGTCT
TTGGACAGTTTGCCTATTTCCAGACAGCCCTGAATGATTTGGCAGAATTATTTGATGGAACCCATGCA
CAGGCCAGACTTCTCAGCTCACTCTCGGGGTCTCACTCAGGGGAAACATTGCCCTTGGCTACGTCAAA
TCAAATTCTGCTCCGATTCAGTGCAAAGAGCGGTGCCTCTGCCCGCGGCTTCCACTTCGTGTATCAAG
CTGTTCCTCGTACCAGTGACACCCAATGCAGCTCTGTCCCCGAGCCCAGATACGGAAGGAGAATTGGT
TCTGAGTTTTCTGCCGGCTCCATCGTCCGATTCGAGTGCAACCCGGGATACCTGCTTCAGGGTTCCAC
GGCGCTCCACTGCCAGTCCGTGCCCAACGCCTTGGCACAGTGGAACGACACGATCCCCAGCTGTGTGG
TACCCTGCAGTGGCAATTTCACTCAACGAAGAGGTACAATCCTGTCCCCCGGCTACCCTGAGCCATAC
GGAAACAACTTGAACTGTATATGGAAGATCATAGTTACGGAGGGCTCGGGAATTCAGATCCAAGTGAT
CAGTTTTGCCACGGAGCAGAACTGGGACTCCCTTGAGATCCACGATGGTGGGGATGTGACCGCACCCA
GACTGGGAAGCTTCTCAGGCACCACAGTACCGGCACTGCTGAACAGTACTTCCAACCAACTCTACCTG
CATTTCCAGTCTGACATTAGTGTGGCAGCTGCTGGTTTCCACCTGGAATACAAAACTGTAGGTCTTGC
TGCATGCCAAGAACCAGCCCTCCCCAGCAACAGCATCAAAATCGGAGATCGGTACATGGTGAACGACG
TGCTCTCCTTCCAGTGCGAGCCCGGGTACACCCTGCAGGGCCGTTCCCACATTTCCTGTATGCCAGGG
ACCGTTCGCCGTTGGAACTATCCGTCTCCCCTGTGCATTGCAACCTGTGGAGGGACGCTGAGCACCTT
GGGTGGTGTGATCCTGAGCCCCGGCTTCCCAGGTTCTTACCCCAACAACTTAGACTGCACCTGGAGGA
TCTCATTACCCATCGGCTATGGTGCACATATTCAGTTTCTGAATTTTTCTACCGAAGCTAATCATGAC
TTCCTTGAAATTCAAAATGGACCTTACCACACCAGCCCCATGATTGGACAATTTAGCGGCACGGATCT
CCCCGCGGCCCTGCTGAGCACAACGCATGAAACCCTCATCCACTTTTATAGTGACCATTCGCAAAACC
GGCAAGGATTTAAACTTGCTTACCAAGCCTATGAATTACAGAACTGTCCAGATCCACCCCCATTTCAG
AATGGGTACATGATCAACTCGGATTACAGCGTGGGGCAATCAGTATCTTTCGAGTGTTATCCTGGGTA
CATTCTAATAGGCCATCCTCCG

SEQ ID NO:2

```
5G-3V1 Nucleotide sequence        6145 bp
   1 TTTTAGGGAT GGTATGAATT TAATATTTTT TAGTATTACA ATATATTCTT
  51 ATAAAAAAGG TCCAAGTGAA AAAGGCGATT GAGTTGAAGT CAAGAGGAGT
 101 CAAGATGCTG CCCAGCAAGG ATGGAAGCCA TAAAAACTCT GTCTGGCATA
 151 TGGAATAACA TCAACCATGT GACATCCGAA GAAGATACGT TCATTATGTA
 201 TCTGGGAAAA CCATGGCTTC AAGTGAAAAT TCAAGTGAGC CAAGGAGGTG
 251 TTGCATTGGT CTCTGACATG TGTCCAGATC CTGGGATTCC AGAAAATGGT
 301 AGAAGAGCAG GTTCCGACTT CAGGGTTGGT GCAAATGTAC AGTTTTCATG
 351 TGAGGACAAT TACGTGCTCC AGGGATCTAA AAGCATCACC TGTCAGAGAG
 401 TTACAGAGAC GCTCGCTGCT TGGAGTGACC ACAGGCCCAT CTGCCGAGCG
 451 AGAACATGTG GATCCAATCT GCGTGGGCCC AGCGGCGTCA TTACCTCCCC
 501 TAATTATCCG GTTCAGTATG AAGATAATGC ACACTGTGTG TGGGTCATCA
 551 CCACCACCGA CCCGGACAAG GTCATCAAGC TTGCCTTNGA AGAGTTTGAG
 601 CTGGAGCGAG GCTATGACAC CCTNACGGTT GGTGATGCTG GGAAGGTGGG
 651 AGACACCAGA TCGGTCTTGT ANGTGCTCAC GGGATCCAGT GTTCCTGACC
 701 TCATTGTGAG CATGAGCAAC CAGATGTGGC TACATCTGCA GTCGGATGAT
 751 AGCATTGGCT CACCTGGGTT TAAAGCTGTT TACCAAGAAA TTGAAAAGGG
 801 AGGGTGTGGG GATCCTGGAA TCCCCGCCTA TGGGAAGCGG ACGGGCAGCA
 851 GTTTCCTCCA TGGAGATACA CTCACCTTTG AATGCCCGGC GGCCTTTGAG
 901 CTGGTGGGGG AGAGAGTTAT CACCTGTCAG CAGAACAATC AGTGGTCTGG
 951 CAACAAGCCC AGCTGTGTAT TTTCATGTTT CTTCAACTTT ACGGCATCAT
1001 CTGGGATTAT TCTGTCACCA AATTATCCAG AGGAATATGG GAACAACATG
1051 AACTGTGTCT GGTTGATTAT CTCGGAGCCA GGAAGTCGAA TTCACCTAAT
1101 CTTTAATGAT TTTGATGTTG AGCCTCAATT TGACTTTCTC GCGGTCAAGG
1151 ATGATGGCAT TTCTGACATA ACTGTCCTGG GTACTTTTTC TGGCAATGAA
1201 GTGCCTTCCC AGCTGGCCAG CAGTGGGCAT ATAGTTCGCT TGGAATTTCA
1251 GTCTGACCAT TCCACTACTG GCAGAGGGTT CAACATCACT TACACCACAT
1301 TTGGTCAGAA TGAGTGCCAT GATCCTGGCA TTCCTATAAA CGGACGACGT
1351 TTTGGTGACA GGTTTCTACT CGGGAGCTCG GTTTCTTTCC ACTGTGATGA
1401 TGGCTTTGTC AAGACCCAGG GATCCGAGTC CATTACCTGC ATACTGCAAG
1451 ACGGGAACGT GGTCTGGAGC TCCACCGTGC CCCGCTGTGA AGCTCCATGT
1501 GGTGGACATC TGACAGCGTC CAGCGGAGTC ATTTTGCCTC CTGGATGGCC
1551 AGGATATTAT AAGGATTCTT TACATTGTGA ATGGATAATT GAAGCAAAAC
1601 CAGGCCACTC TATCAAAATA ACTTTTGACA GATTTCAGAC AGAGGTCAAT
1651 TATGACACCT TGGAGGTCAG AGATGGGCCA GCCAGTTCGT CCCCACTGAT
1701 CGGCGAGTAC CACGGCACCC AGGCACCCCA GTTCCTCATC AGCACCGGGA
1751 ACTTCATGTA CCTGCTATTC ACCACTGACA ACAGCCGCTC CAGCATCGGC
1801 TTCCTCATCC ACTATGAGAG TGTGACGCTT GAGTCGGATT CCTGCCTGGA
1851 CCCGGGCATC CCTGTGAACG GCCATCGCCA CGGTGGAGAC TTTGGCATCA
1901 GGTCCACAGT GACTTTCAGC TGTGACCCGG GGTACACACT AAGTGACGAC
1951 GAGCCCCTCG TCTGTGAGAG GAACCACCAG TGGAACCACG CCTTGCCCAG
2001 CTGCGACGCT CTATGTGGAG GCTACATCCA AGGGAAGAGT GGAACAGTCC
2051 TTTCTCCTGG GTTTCCAGAT TTTTATCCAA ACTCTCTAAA CTGCACGTGG
2101 ACCATTGAAG TGTCTCATGG GAAAGGAGTT CAAATGATCT TTCACACCTT
2151 TCATCTTGAG AGTTCCCACG ACTATTTACT GATCACAGAG GATGGAAGTT
2201 TTTCCGAGCC CGTTGCCAGG CTCACCGGGT CGGTGTTGCC TCATACGATC
2251 AAGGCAGGCC TGTTNGGAAA CTTCACTGCC CAGCTTCGGT TTATATCAGA
2301 CTTCTCAATT TCGTACGAGG GCTTCAATAT CACATTTTCA GAATATGACC
2351 TGGAGCCATG TGATGATCCT GGAGTCCCTG CCTTCAGCCG AAGAATTGGT
2401 TTTCACTTTG GTGTGGGAGA CTCTCTGACG TTTTCCTGCT TCCTGGGATA
2451 TCGTTTAGAA GGTGCCACCA AGCTTACCTG CCTGGGTGGG GGCCGCCGTG
2501 TGTGGAGTGC ACCTCTGCCA AGGTGTGTGG CCGAATGTGG AGCAAGTGTC
2551 AAAGGAAATG AAGGAACATT ACTGTCTCCA AATTTTCCAT CCAATTATGA
2601 TAATAACCAT GAGTGTATCT ATAAAATAGA AACAGAAGCC GGCAAGGGCA
2651 TCCACCTTAG AACACGAAGC TTCCAGCTGT TTGAAGGAGA TACTCTAAAG
2701 GTATATGATG GAAAAGACAG TTCCTCACGT CCACTGGGCA CGTTCACTAA
2751 AAATGAACTT CTGGGGCTGA TCCTAAACAG CACATCCAAT CACCTGTGGC
2801 TAGAGTTCAA CACCAATGGA TCTGACACCG ACCAAGGTTT TCAACTCACC
2851 TATACCAGTT TTGATCTGGT AAAATGTGAG GATCCGGGCA TCCCTAACTA
2901 CGGCTATAGG ATCCGTGATG AAGGCCACTT TACCGACACT GTAGTTCTGT
2951 ACAGTTGCAA CCCGGGGTAC GCCATGCATG GCAGCAACAC CCTGACCTGT
3001 TTGAGTGGAG ACAGGAGAGT GTGGGACAAA CCACTACCTT CGTGCATAGC
```

```
3051  GGAATGTGGT GGTCAGATCC ATGCAGCCAC ATCAGGACGA ATATTGTCCC
3101  CTGGCTATCC AGCTCCGTAT GACAACAACC TCCACTGCAC CTGGATTATA
3151  GAGGCAGACC CAGGAAAGAC CATTAGCCTC CATTTCATTG TTTTCGACAC
3201  GGAGATGGCT CACGACATCC TCAAGGTCTG GGACGGGCCG GTGGACAGTG
3251  ACATCCTGCT GAAGGAGTGG AGTGGCTCCG CCCTTCCGGA GGACATCCAC
3301  AGCACCTTCA ACTCACTCAC CCTGCAGTTC GACAGCGACT TCTTCATCAG
3351  CAAGTCTGGC TTCTCCATCC AGTTCTCCAC CTCAATTGCA GCCACCTGTA
3401  ACGATCCAGG TATGCCCCAA AATGGCACCC GCTATGGAGA CAGCAGAGAG
3451  GCTGGAGACA CCGTCACATT CCAGTGTGAC CCTGGCTATC AGCTCCAAGG
3501  ACAAGCCAAA ATCACCTGTG TGCAGCTGAA TAACCGGTTC TTTTGGCAAC
3551  CAGACCCTCC TACATGCATA GCTGCTTGTG GAGGGAATCT GACGGGCCCA
3601  GCAGGTGTTA TTTTGTCACC CAACTACCCA CAGCCGTATC CTCCTGGGAA
3651  GGAATGTGAC TGGAGAGTAA AAGTGAACCC GGACTTTGTC ATCGCCTTGA
3701  TATTCAAAAG TTTCAACATG GAGCCCAGCT ATGACTTCCT ACACATCTAT
3751  GAAGGGGAAG ATTCCAACAG CCCCCTCATT GGGAGTTACC AGGGCTCTCA
3801  GGCCCCAGAA AGAATAGAGA GTAGCGGAAA CAGCCTGTTT CTGGCATTTC
3851  GGAGTGATGC CTCCGTGGGC CTTTCAGGGT TCGCCATTGA ATTTAAAGAG
3901  AAACCACGGG AAGCTTGTTT TGACCCAGGA AATATAATGA ATGGGACAAG
3951  AGTTGGAACA GACTTCAAGC TTGGCTCCAC CATCACCTAC CAGTGTGACT
4001  CTGGCTATAA GATTCTTGAC CCCTCATCCA TCACCTGTGT GATTGGGGCT
4051  GATGGGAAAC CCTCCTGGGA CCAAGTGCTG CCCTCCTGCA ATGCTCCCTG
4101  TGGAGGCCAG TACACGGGAT CAGAAGGGGT AGTTTTATCA CCAAACTACC
4151  CCCATAATTA CACAGCTGGT CAAATATGCC TCTATTCCAT CACGGTACCA
4201  AAGGAATTCG TGGTCTTTGG ACAGTTTGCC TATTTCCAGA CAGCCCTGAA
4251  TGATTTGGCA GAATTATTTG ATGGAACCCA TGCACAGGCC AGACTTCTCA
4301  GCTCACTCTC GGGGTCTCAC TCAGGGGAAA CATTGCCCTT GGCTACGTCA
4351  AATCAAATTC TGCTCCGATT CAGTGCAAAG AGCGGTGCCT CTGCCCGCGG
4401  CTTCCACTTC GTGTATCAAG CTGTTCCTCG TACCAGTGAC ACCCAATGCA
4451  GCTCTGTCCC CGAGCCCAGA TACGGAAGGA GAATTGGTTC TGAGTTTTCT
4501  GCCGGCTCCA TCGTCCGATT CGAGTGCAAC CCGGGATACC TGCTTCAGGG
4551  TTCCACGGCG CTCCACTGCC AGTCCGTGCC CAACGCCTTG GCACAGTGGA
4601  ACGACACGAT CCCCAGCTGT GTGGTACCCT GCAGTGGCAA TTTCACTCAA
4651  CGAAGAGGTA CAATCCTGTC CCCCGGCTAC CCTGAGCCAT ACGGAAACAA
4701  CTTGAACTGT ATATGGAAGA TCATAGTTAC GGAGGGCTCG GGAATTCAGA
4751  TCCAAGTGAT CAGTTTTGCC ACGGAGCAGA ACTGGGACTC CCTTGAGATC
4801  CACGATGGTG GGGATGTGAC CGCACCCAGA CTGGGAAGCT TCTCAGGCAC
4851  CACAGTACCG GCACTGCTGA ACAGTACTTC CAACCAACTC TACCTGCATT
4901  TCCAGTCTGA CATTAGTGTG GCAGCTGCTG GTTTCCACCT GGAATACAAA
4951  ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGCAT
5001  CAAAATCGGA GATCGGTACA TGGTGAACGA CGTGCTCTCC TTCCAGTGCG
5051  AGCCCGGGTA CACCCTGCAG GGCCGTTCCC ACATTTCCTG TATGCCAGGG
5101  ACCGTTCGCC GTTGGAACTA TCCGTCTCCC CTGTGCATTG CAACCTGTGG
5151  AGGGACGCTG AGCACCTTGG GTGGTGTGAT CCTGAGCCCC GGCTTCCCAG
5201  GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC
5251  GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA
5301  TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG
5351  GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT
5401  GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT
5451  TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT
5501  CTAAATACAC TTCTTACATG TAAATTGTAT TTAAGTATAA ATCTCCCTAA
5551  CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT
5601  TCTGGTTAGT AGTGGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA
5651  CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG
5701  ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC
5751  AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG
5801  CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG
5851  CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG
5901  AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT
5951  ACTAAAAATA CCAAAAATTA GCCGGGCATA GTGGCGGGCG CCTGTAGTCC
6001  CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG
6051  AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG
6101  AGCCAGACTC CATCTCGAAA AAAAAAAAAA AAAAAAAAAA AAAAA
```

SEQ ID NO:3

```
5G-3V2 Nucleotide sequence        6409 bp
   1 TTTTAGGGAT GGTATGAATT TAATATTTTT TAGTATTACA ATATATTCTT
  51 ATAAAAAAGG TCCAAGTGAA AAAGGCGATT GAGTTGAAGT CAAGAGGAGT
 101 CAAGATGCTG CCCAGCAAGG ATGGAAGCCA TAAAAACTCT GTCTGGCATA
 151 TGGAATAACA TCAACCATGT GACATCCGAA GAAGATACGT TCATTATGTA
 201 TCTGGGAAAA CCATGGCTTC AAGTGAAAAT TCAAGTGAGC CAAGGAGGTG
 251 TTGCATTGGT CTCTGACATG TGTCCAGATC CTGGGATTCC AGAAAATGGT
 301 AGAAGAGCAG GTTCCGACTT CAGGGTTGGT GCAAATGTAC AGTTTTCATG
 351 TGAGGACAAT TACGTGCTCC AGGGATCTAA AAGCATCACC TGTCAGAGAG
 401 TTACAGAGAC GCTCGCTGCT TGGAGTGACC ACAGGCCCAT CTGCCGAGCG
 451 AGAACATGTG GATCCAATCT GCGTGGGCCC AGCGGCGTCA TTACCTCCCC
 501 TAATTATCCG GTTCAGTATG AAGATAATGC ACACTGTGTG TGGGTCATCA
 551 CCACCACCGA CCCGGACAAG GTCATCAAGC TTGCCTTNGA AGAGTTTGAG
 601 CTGGAGCGAG GCTATGACAC CCTNACGGTT GGTGATGCTG GGAAGGTGGG
 651 AGACACCAGA TCGGTCTTGT ANGTGCTCAC GGGATCCAGT GTTCCTGACC
 701 TCATTGTGAG CATGAGCAAC CAGATGTGGC TACATCTGCA GTCGGATGAT
 751 AGCATTGGCT CACCTGGGTT TAAAGCTGTT TACCAAGAAA TTGAAAAGGG
 801 AGGGTGTGGG GATCCTGGAA TCCCCGCCTA TGGGAAGCGG ACGGGCAGCA
 851 GTTTCCTCCA TGGAGATACA CTCACCTTTG AATGCCCGGC GGCCTTTGAG
 901 CTGGTGGGGG AGAGAGTTAT CACCTGTCAG CAGAACAATC AGTGGTCTGG
 951 CAACAAGCCC AGCTGTGTAT TTTCATGTTT CTTCAACTTT ACGGCATCAT
1001 CTGGGATTAT TCTGTCACCA AATTATCCAG AGGAATATGG GAACAACATG
1051 AACTGTGTCT GGTTGATTAT CTCGGAGCCA GGAAGTCGAA TTCACCTAAT
1101 CTTTAATGAT TTTGATGTTG AGCCTCAATT TGACTTTCTC GCGGTCAAGG
1151 ATGATGGCAT TTCTGACATA ACTGTCCTGG GTACTTTTTC TGGCAATGAA
1201 GTGCCTTCCC AGCTGGCCAG CAGTGGGCAT ATAGTTCGCT TGGAATTTCA
1251 GTCTGACCAT TCCACTACTG GCAGAGGGTT CAACATCACT TACACCACAT
1301 TTGGTCAGAA TGAGTGCCAT GATCCTGGCA TTCCTATAAA CGGACGACGT
1351 TTTGGTGACA GGTTTCTACT CGGGAGCTCG GTTTCTTTCC ACTGTGATGA
1401 TGGCTTTGTC AAGACCCAGG GATCCGAGTC CATTACCTGC ATACTGCAAG
1451 ACGGGAACGT GGTCTGGAGC TCCACCGTGC CCCGCTGTGA AGCTCCATGT
1501 GGTGGACATC TGACAGCGTC CAGCGGAGTC ATTTTGCCTC CTGGATGGCC
1551 AGGATATTAT AAGGATTCTT TACATTGTGA ATGGATAATT GAAGCAAAAC
1601 CAGGCCACTC TATCAAAATA ACTTTTGACA GATTTCAGAC AGAGGTCAAT
1651 TATGACACCT TGGAGGTCAG AGATGGGCCA GCCAGTTCGT CCCCACTGAT
1701 CGGCGAGTAC CACGGCACCC AGGCACCCCA GTTCCTCATC AGCACCGGGA
1751 ACTTCATGTA CCTGCTATTC ACCACTGACA ACAGCCGCTC CAGCATCGGC
1801 TTCCTCATCC ACTATGAGAG TGTGACGCTT GAGTCGGATT CCTGCCTGGA
1851 CCCGGGCATC CCTGTGAACG GCCATCGCCA CGGTGGAGAC TTTGGCATCA
1901 GGTCCACAGT GACTTTCAGC TGTGACCCGG GGTACACACT AAGTGACGAC
1951 GAGCCCCTCG TCTGTGAGAG GAACCACCAG TGGAACCACG CCTTGCCCAG
2001 CTGCGACGCT CTATGTGGAG GCTACATCCA AGGGAAGAGT GGAACAGTCC
2051 TTTCTCCTGG GTTTCCAGAT TTTTATCCAA ACTCTCTAAA CTGCACGTGG
2101 ACCATTGAAG TGTCTCATGG GAAAGGAGTT CAAATGATCT TTCACACCTT
2151 TCATCTTGAG AGTTCCCACG ACTATTTACT GATCACAGAG GATGGAAGTT
2201 TTTCCGAGCC CGTTGCCAGG CTCACCGGGT CGGTGTTGCC TCATACGATC
2251 AAGGCAGGCC TGTTNGGAAA CTTCACTGCC CAGCTTCGGT TTATATCAGA
2301 CTTCTCAATT TCGTACGAGG GCTTCAATAT CACATTTTCA GAATATGACC
2351 TGGAGCCATG TGATGATCCT GGAGTCCCTG CCTTCAGCCG AAGAATTGGT
2401 TTTCACTTTG GTGTGGGAGA CTCTCTGACG TTTTCCTGCT TCCTGGGATA
2451 TCGTTTAGAA GGTGCCACCA AGCTTACCTG CCTGGGTGGG GGCCGCCGTG
2501 TGTGGAGTGC ACCTCTGCCA AGGTGTGTGG CCGAATGTGG AGCAAGTGTC
2551 AAAGGAAATG AAGGAACATT ACTGTCTCCA AATTTTCCAT CCAATTATGA
2601 TAATAACCAT GAGTGTATCT ATAAAATAGA AACAGAAGCC GGCAAGGGCA
2651 TCCACCTTAG AACACGAAGC TTCCAGCTGT TTGAAGGAGA TACTCTAAAG
2701 GTATATGATG GAAAAGACAG TTCCTCACGT CCACTGGGCA CGTTCACTAA
2751 AAATGAACTT CTGGGGCTGA TCCTAAACAG CACATCCAAT CACCTGTGGC
2801 TAGAGTTCAA CACCAATGGA TCTGACACCG ACCAAGGTTT TCAACTCACC
2851 TATACCAGTT TTGATCTGGT AAAATGTGAG GATCCGGGCA TCCCTAACTA
2901 CGGCTATAGG ATCCGTGATG AAGGCCACTT TACCGACACT GTAGTTCTGT
2951 ACAGTTGCAA CCCGGGGTAC GCCATGCATG GCAGCAACAC CCTGACCTGT
```

```
3001 TTGAGTGGAG ACAGGAGAGT GTGGGACAAA CCACTACCTT CGTGCATAGC
3051 GGAATGTGGT GGTCAGATCC ATGCAGCCAC ATCAGGACGA ATATTGTCCC
3101 CTGGCTATCC AGCTCCGTAT GACAACAACC TCCACTGCAC CTGGATTATA
3151 GAGGCAGACC CAGGAAAGAC CATTAGCCTC CATTTCATTG TTTTCGACAC
3201 GGAGATGGCT CACGACATCC TCAAGGTCTG GGACGGGCCG GTGGACAGTG
3251 ACATCCTGCT GAAGGAGTGG AGTGGCTCCG CCCTTCCGGA GGACATCCAC
3301 AGCACCTTCA ACTCACTCAC CCTGCAGTTC GACAGCGACT TCTTCATCAG
3351 CAAGTCTGGC TTCTCCATCC AGTTCTCCAC CTCAATTGCA GCCACCTGTA
3401 ACGATCCAGG TATGCCCCAA AATGGCACCC GCTATGGAGA CAGCAGAGAG
3451 GCTGGAGACA CCGTCACATT CCAGTGTGAC CCTGGCTATC AGCTCCAAGG
3501 ACAAGCCAAA ATCACCTGTG TGCAGCTGAA TAACCGGTTC TTTTGGCAAC
3551 CAGACCCTCC TACATGCATA GCTGCTTGTG GAGGGAATCT GACGGGCCCA
3601 GCAGGTGTTA TTTTGTCACC CAACTACCCA CAGCCGTATC CTCCTGGGAA
3651 GGAATGTGAC TGGAGAGTAA AAGTGAACCC GGACTTTGTC ATCGCCTTGA
3701 TATTCAAAAG TTTCAACATG GAGCCCAGCT ATGACTTCCT ACACATCTAT
3751 GAAGGGGAAG ATTCCAACAG CCCCCTCATT GGGAGTTACC AGGGCTCTCA
3801 GGCCCCAGAA AGAATAGAGA GTAGCGGAAA CAGCCTGTTT CTGGCATTTC
3851 GGAGTGATGC CTCCGTGGGC CTTTCAGGGT TCGCCATTGA ATTTAAAGAG
3901 AAACCACGGG AAGCTTGTTT TGACCCAGGA AATATAATGA ATGGGACAAG
3951 AGTTGGAACA GACTTCAAGC TTGGCTCCAC CATCACCTAC CAGTGTGACT
4001 CTGGCTATAA GATTCTTGAC CCCTCATCCA TCACCTGTGT GATTGGGGCT
4051 GATGGGAAAC CCTCCTGGGA CCAAGTGCTG CCCTCCTGCA ATGCTCCCTG
4101 TGGAGGCCAG TACACGGGAT CAGAAGGGGT AGTTTTATCA CCAAACTACC
4151 CCCATAATTA CACAGCTGGT CAAATATGCC TCTATTCCAT CACGGTACCA
4201 AAGGAATTCG TGGTCTTTGG ACAGTTTGCC TATTTCCAGA CAGCCCTGAA
4251 TGATTTGGCA GAATTATTTG ATGGAACCCA TGCACAGGCC AGACTTCTCA
4301 GCTCACTCTC GGGGTCTCAC TCAGGGGAAA CATTGCCCTT GGCTACGTCA
4351 AATCAAATTC TGCTCCGATT CAGTGCAAAG AGCGGTGCCT CTGCCCGCGG
4401 CTTCCACTTC GTGTATCAAG CTGTTCCTCG TACCAGTGAC ACCCAATGCA
4451 GCTCTGTCCC CGAGCCCAGA TACGGAAGGA GAATTGGTTC TGAGTTTTCT
4501 GCCGGCTCCA TCGTCCGATT CGAGTGCAAC CCGGGATACC TGCTTCAGGG
4551 TTCCACGGCG CTCCACTGCC AGTCCGTGCC CAACGCCTTG GCACAGTGGA
4601 ACGACACGAT CCCCAGCTGT GTGGTACCCT GCAGTGGCAA TTTCACTCAA
4651 CGAAGAGGTA CAATCCTGTC CCCCGGCTAC CCTGAGCCAT ACGGAAACAA
4701 CTTGAACTGT ATATGGAAGA TCATAGTTAC GGAGGGCTCG GGAATTCAGA
4751 TCCAAGTGAT CAGTTTTGCC ACGGAGCAGA ACTGGGACTC CCTTGAGATC
4801 CACGATGGTG GGGATGTGAC CGCACCCAGA CTGGGAAGCT TCTCAGGCAC
4851 CACAGTACCG GCACTGCTGA ACAGTACTTC CAACCAACTC TACCTGCATT
4901 TCCAGTCTGA CATTAGTGTG GCAGCTGCTG GTTTCCACCT GGAATACAAA
4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGCAT
5001 CAAAATCGGA GATCGGTACA TGGTGAACGA CGTGCTCTCC TTCCAGTGCG
5051 AGCCCGGGTA CACCCTGCAG GGCCGTTCCC ACATTTCCTG TATGCCAGGG
5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCCC CTGTGCATTG CAACCTGTGG
5151 AGGGACGCTG AGCACCTTGG GTGGTGTGAT CCTGAGCCCC GGCTTCCCAG
5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC
5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA
5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG
5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT
5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT
5451 TAAACTTGCT TACCAAGCCT ATGAATTACA GAACTGTCCA GATCCACCCC
5501 CATTTCAGAA TGGGTACATG ATCAACTCGG ATTACAGCGT GGGGCAATCA
5551 GTATCTTTCG AGTGTTATCC TGGGTACATT CTAATAGGCC ATCCTGTCCT
5601 CACTTGTCAG CATGGGATCA ACAGAAACTG GAACTACCCT TTTCCAAGAT
5651 GTGATGCCCC TTGTGGGTAC AACGTAACTT CTCAGAACGG CACCATCTAC
5701 TCCCCTGGCT TTCCTGATGA GTATCCGATC CTGAAGGACT GCATTTGGCT
5751 CATCACGGTG CCTCCAGGGC ACGGAGTTTA CATCAACTTC ACCCTGTTAC
5801 AGACGGAAGC TGTCAACGAT TACATTGCTG TTTGGGACGG TCCCGATCAG
5851 AACTCACCCC AGCTGGGAGT TTTCAGTGGC AACACAGCCC TCGAAACGGC
5901 GTATAGCTCC ACCAACCAAG TCCTGCTCAA GTTCCACAGC GACTTTTCAA
5951 ATGGAGGCTT CTTTGTCCTC AATTTCCACG GTCAGTTGAT TTTCACTCCG
6001 TTAGTTAAGA CTGAGAATTC CATGTGGTGT TTACTGCAGT GTTGTCCCAC
6051 GCCTTGTTTC CAGCTGAAGT TTCTTGATTC AGCCGAGGGC GTGTATGATT
6101 CTTTTGCACT GGAGGCCAGC GTTTCCTGTG GTCCTTTTTT TGTTTAATGA
```

```
6151   TGTCTTTATT ATTTCACATC GTATCCAGCT TGGATTTATT CCAAGATACA
6201   TGTATCCTAA GTGAAACTCT AAGATGAAGA CCATTGAAAG AGATTTGGTA
6251   CCTTTTATAG ATTTACTCAT CCCTGTCTCA AGATAAGGTG TTATAGCAAA
6301   TGTCATGTAA CTATAAATGG TGTGAAAGCA AACCTCCAAT AATCCTGGGA
6351   ATGCACTCTA AACGATATGT AGAACATCTG TCAATCNATC GCTTATCTCT
6401   CACGAACAC
```

SEQ ID NO:4

```
5G-3V3 Nucleotide sequence        5667 bp
   1 TTTTAGGGAT GGTATGAATT TAATATTTTT TAGTATTACA ATATATTCTT
  51 ATAAAAAAGG TCCAAGTGAA AAAGGCGATT GAGTTGAAGT CAAGAGGAGT
 101 CAAGATGCTG CCCAGCAAGG ATGGAAGCCA TAAAAACTCT GTCTGGCATA
 151 TGGAATAACA TCAACCATGT GACATCCGAA GAAGATACGT TCATTATGTA
 201 TCTGGGAAAA CCATGGCTTC AAGTGAAAAT TCAAGTGAGC CAAGGAGGTG
 251 TTGCATTGGT CTCTGACATG TGTCCAGATC CTGGGATTCC AGAAAATGGT
 301 AGAAGAGCAG GTTCCGACTT CAGGGTTGGT GCAAATGTAC AGTTTTCATG
 351 TGAGGACAAT TACGTGCTCC AGGGATCTAA AAGCATCACC TGTCAGAGAG
 401 TTACAGAGAC GCTCGCTGCT TGGAGTGACC ACAGGCCCAT CTGCCGAGCG
 451 AGAACATGTG GATCCAATCT GCGTGGGCCC AGCGGCGTCA TTACCTCCCC
 501 TAATTATCCG GTTCAGTATG AAGATAATGC ACACTGTGTG TGGGTCATCA
 551 CCACCACCGA CCCGGACAAG GTCATCAAGC TTGCCTTNGA AGAGTTTGAG
 601 CTGGAGCGAG GCTATGACAC CCTNACGGTT GGTGATGCTG GGAAGGTGGG
 651 AGACACCAGA TCGGTCTTGT ANGTGCTCAC GGGATCCAGT GTTCCTGACC
 701 TCATTGTGAG CATGAGCAAC CAGATGTGGC TACATCTGCA GTCGGATGAT
 751 AGCATTGGCT CACCTGGGTT TAAAGCTGTT TACCAAGAAA TTGAAAAGGG
 801 AGGGTGTGGG GATCCTGGAA TCCCCGCCTA TGGGAAGCGG ACGGGCAGCA
 851 GTTTCCTCCA TGGAGATACA CTCACCTTTG AATGCCCGGC GGCCTTTGAG
 901 CTGGTGGGGG AGAGAGTTAT CACCTGTCAG CAGAACAATC AGTGGTCTGG
 951 CAACAAGCCC AGCTGTGTAT TTTCATGTTT CTTCAACTTT ACGGCATCAT
1001 CTGGGATTAT TCTGTCACCA AATTATCCAG AGGAATATGG GAACAACATG
1051 AACTGTGTCT GGTTGATTAT CTCGGAGCCA GGAAGTCGAA TTCACCTAAT
1101 CTTTAATGAT TTTGATGTTG AGCCTCAATT TGACTTTCTC GCGGTCAAGG
1151 ATGATGGCAT TTCTGACATA ACTGTCCTGG GTACTTTTTC TGGCAATGAA
1201 GTGCCTTCCC AGCTGGCCAG CAGTGGGCAT ATAGTTCGCT TGGAATTTCA
1251 GTCTGACCAT TCCACTACTG GCAGAGGGTT CAACATCACT TACACCACAT
1301 TTGGTCAGAA TGAGTGCCAT GATCCTGGCA TTCCTATAAA CGGACGACGT
1351 TTTGGTGACA GGTTTCTACT CGGGAGCTCG GTTTCTTTCC ACTGTGATGA
1401 TGGCTTTGTC AAGACCCAGG GATCCGAGTC CATTACCTGC ATACTGCAAG
1451 ACGGGAACGT GGTCTGGAGC TCCACCGTGC CCCGCTGTGA AGCTCCATGT
1501 GGTGGACATC TGACAGCGTC CAGCGGAGTC ATTTTGCCTC CTGGATGGCC
1551 AGGATATTAT AAGGATTCTT TACATTGTGA ATGGATAATT GAAGCAAAAC
1601 CAGGCCACTC TATCAAAATA ACTTTTGACA GATTTCAGAC AGAGGTCAAT
1651 TATGACACCT TGGAGGTCAG AGATGGGCCA GCCAGTTCGT CCCCACTGAT
1701 CGGCGAGTAC CACGGCACCC AGGCACCCCA GTTCCTCATC AGCACCGGGA
1751 ACTTCATGTA CCTGCTATTC ACCACTGACA ACAGCCGCTC CAGCATCGGC
1801 TTCCTCATCC ACTATGAGAG TGTGACGCTT GAGTCGGATT CCTGCCTGGA
1851 CCCGGGCATC CCTGTGAACG GCCATCGCCA CGGTGGAGAC TTTGGCATCA
1901 GGTCCACAGT GACTTTCAGC TGTGACCCGG GGTACACACT AAGTGACGAC
1951 GAGCCCCTCG TCTGTGAGAG GAACCACCAG TGGAACCACG CCTTGCCCAG
2001 CTGCGACGCT CTATGTGGAG GCTACATCCA AGGGAAGAGT GGAACAGTCC
2051 TTTCTCCTGG GTTTCCAGAT TTTTATCCAA ACTCTCTAAA CTGCACGTGG
2101 ACCATTGAAG TGTCTCATGG GAAAGGAGTT CAAATGATCT TTCACACCTT
2151 TCATCTTGAG AGTTCCCACG ACTATTTACT GATCACAGAG GATGGAAGTT
2201 TTTCCGAGCC CGTTGCCAGG CTCACCGGGT CGGTGTTGCC TCATACGATC
2251 AAGGCAGGCC TGTTNGGAAA CTTCACTGCC CAGCTTCGGT TTATATCAGA
2301 CTTCTCAATT TCGTACGAGG GCTTCAATAT CACATTTTCA GAATATGACC
2351 TGGAGCCATG TGATGATCCT GGAGTCCCTG CCTTCAGCCG AAGAATTGGT
2401 TTTCACTTTG GTGTGGGAGA CTCTCTGACG TTTTCCTGCT TCCTGGGATA
2451 TCGTTTAGAA GGTGCCACCA AGCTTACCTG CCTGGGTGGG GGCCGCCGTG
2501 TGTGGAGTGC ACCTCTGCCA AGGTGTGTGG CCGAATGTGG AGCAAGTGTC
2551 AAAGGAAATG AAGGAACATT ACTGTCTCCA AATTTTCCAT CCAATTATGA
2601 TAATAACCAT GAGTGTATCT ATAAAATAGA AACAGAAGCC GGCAAGGGCA
2651 TCCACCTTAG AACACGAAGC TTCCAGCTGT TTGAAGGAGA TACTCTAAAG
2701 GTATATGATG GAAAAGACAG TTCCTCACGT CCACTGGGCA CGTTCACTAA
2751 AAATGAACTT CTGGGGCTGA TCCTAAACAG CACATCCAAT CACCTGTGGC
2801 TAGAGTTCAA CACCAATGGA TCTGACACCG ACCAAGGTTT TCAACTCACC
2851 TATACCAGTT TTGATCTGGT AAAATGTGAG GATCCGGGCA TCCCTAACTA
2901 CGGCTATAGG ATCCGTGATG AAGGCCACTT TACCGACACT GTAGTTCTGT
2951 ACAGTTGCAA CCCGGGGTAC GCCATGCATG GCAGCAACAC CCTGACCTGT
```

```
3001 TTGAGTGGAG ACAGGAGAGT GTGGGACAAA CCACTACCTT CGTGCATAGC
3051 GGAATGTGGT GGTCAGATCC ATGCAGCCAC ATCAGGACGA ATATTGTCCC
3101 CTGGCTATCC AGCTCCGTAT GACAACAACC TCCACTGCAC CTGGATTATA
3151 GAGGCAGACC CAGGAAAGAC CATTAGCCTC CATTTCATTG TTTTCGACAC
3201 GGAGATGGCT CACGACATCC TCAAGGTCTG GGACGGGCCG GTGGACAGTG
3251 ACATCCTGCT GAAGGAGTGG AGTGGCTCCG CCCTTCCGGA GGACATCCAC
3301 AGCACCTTCA ACTCACTCAC CCTGCAGTTC GACAGCGACT TCTTCATCAG
3351 CAAGTCTGGC TTCTCCATCC AGTTCTCCAC CTCAATTGCA GCCACCTGTA
3401 ACGATCCAGG TATGCCCCAA AATGGCACCC GCTATGGAGA CAGCAGAGAG
3451 GCTGGAGACA CCGTCACATT CCAGTGTGAC CCTGGCTATC AGCTCCAAGG
3501 ACAAGCCAAA ATCACCTGTG TGCAGCTGAA TAACCGGTTC TTTTGGCAAC
3551 CAGACCCTCC TACATGCATA GCTGCTTGTG GAGGGAATCT GACGGGCCCA
3601 GCAGGTGTTA TTTTGTCACC CAACTACCCA CAGCCGTATC CTCCTGGGAA
3651 GGAATGTGAC TGGAGAGTAA AAGTGAACCC GGACTTTGTC ATCGCCTTGA
3701 TATTCAAAAG TTTCAACATG GAGCCCAGCT ATGACTTCCT ACACATCTAT
3751 GAAGGGGAAG ATTCCAACAG CCCCCTCATT GGGAGTTACC AGGGCTCTCA
3801 GGCCCCAGAA AGAATAGAGA GTAGCGGAAA CAGCCTGTTT CTGGCATTTC
3851 GGAGTGATGC CTCCGTGGGC CTTTCAGGGT TCGCCATTGA ATTTAAAGAG
3901 AAACCACGGG AAGCTTGTTT TGACCCAGGA AATATAATGA ATGGGACAAG
3951 AGTTGGAACA GACTTCAAGC TTGGCTCCAC CATCACCTAC CAGTGTGACT
4001 CTGGCTATAA GATTCTTGAC CCCTCATCCA TCACCTGTGT GATTGGGGCT
4051 GATGGGAAAC CCTCCTGGGA CCAAGTGCTG CCCTCCTGCA ATGCTCCCTG
4101 TGGAGGCCAG TACACGGGAT CAGAAGGGGT AGTTTTATCA CCAAACTACC
4151 CCCATAATTA CACAGCTGGT CAAATATGCC TCTATTCCAT CACGGTACCA
4201 AAGGAATTCG TGGTCTTTGG ACAGTTTGCC TATTTCCAGA CAGCCCTGAA
4251 TGATTTGGCA GAATTATTTG ATGGAACCCA TGCACAGGCC AGACTTCTCA
4301 GCTCACTCTC GGGGTCTCAC TCAGGGGAAA CATTGCCCTT GGCTACGTCA
4351 AATCAAATTC TGCTCCGATT CAGTGCAAAG AGCGGTGCCT CTGCCCGCGG
4401 CTTCCACTTC GTGTATCAAG CTGTTCCTCG TACCAGTGAC ACCCAATGCA
4451 GCTCTGTCCC CGAGCCCAGA TACGGAAGGA GAATTGGTTC TGAGTTTTCT
4501 GCCGGCTCCA TCGTCCGATT CGAGTGCAAC CCGGGATACC TGCTTCAGGG
4551 TTCCACGGCG CTCCACTGCC AGTCCGTGCC CAACGCCTTG GCACAGTGGA
4601 ACGACACGAT CCCCAGCTGT GTGGTACCCT GCAGTGGCAA TTTCACTCAA
4651 CGAAGAGGTA CAATCCTGTC CCCCGGCTAC CCTGAGCCAT ACGGAAACAA
4701 CTTGAACTGT ATATGGAAGA TCATAGTTAC GGAGGGCTCG GGAATTCAGA
4751 TCCAAGTGAT CAGTTTTGCC ACGGAGCAGA ACTGGGACTC CCTTGAGATC
4801 CACGATGGTG GGGATGTGAC CGCACCCAGA CTGGGAAGCT TCTCAGGCAC
4851 CACAGTACCG GCACTGCTGA ACAGTACTTC CAACCAACTC TACCTGCATT
4901 TCCAGTCTGA CATTAGTGTG GCAGCTGCTG GTTTCCACCT GGAATACAAA
4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGCAT
5001 CAAAATCGGA GATCGGTACA TGGTGAACGA CGTGCTCTCC TTCCAGTGCG
5051 AGCCCGGGTA CACCCTGCAG GGCCGTTCCC ACATTTCCTG TATGCCAGGG
5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCCC CTGTGCATTG CAACCTGTGG
5151 AGGGACGCTG AGCACCTTGG GTGGTGTGAT CCTGAGCCCC GGCTTCCCAG
5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC
5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA
5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG
5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT
5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT
5451 TAAACTTGCT TACCAAGCCT AATCTGGAAA CATTGGTCCT GCTTTCCCAT
5501 GTCTTGACAC CCCATTCCAA GCCAGATGTC AAGGAGAAGA AAGGACTTTC
5551 AATTAAAAAA AAAACAAAAA CTCGAAACAA CATGTTTTTT ATTGTACGCC
5601 ATTAATTTCC TATCACTGAG ATATAAAAAT AAATAATGCC NAAAAAAAAA
5651 AAAAAAAAAA AAAAAAA
```

SEQ ID NO:5

5R-3V2 Nucleotide sequence 7323 bp

```
   1 GCGTCGGATG CGCGGCGGGT CTTGGGACCG GGCNCTCTCT CCGGCTCGCC
  51 TTGCCCTCGG GTGATTATTT GGCTCCGCTC ATAGCCCTGC CTTCCTCGGA
 101 GGAGCCATCG GTGTCGCGTG CGTGTGGNGT ATCTGCAGAC ATGACTGCGT
 151 GGAGGAGATT CCAGTCGCTG CTCCTGCTTC TCGGGCTGCT GGTGCTGTGC
 201 GCGAGGCTCC TCACTGCAGC GAAGGGTCAG AACTGTGGAG GCTTAGTCCA
 251 GGGTCCCAAT GGCACTATTG AGAGCCCAGG GTTTCCTCAC GGGTATCCGA
 301 ACTATGCCAA CTGCACCTGG ATCATCATCA CGGGCGAGCG CAATAGGATA
 351 CAGTTGTCCT TCCATACCTT TGCTCTTGAA GAAGATTTTG ATATTTTATC
 401 AGTTTACGAT GGACAGCCTC AACAAGGGAA TTTAAAAGTG AGATTATCGG
 451 GATTTCAGCT GCCCTCCTCT ATAGTGAGTA CAGGATCTAT CCTCACTCTG
 501 TGGTTCACGA CAGACTTCGC TGTGAGTGCC CAAGGTTTCA AAGCATTATA
 551 TGAAGTTTTA CCTAGCCACA CTTGTGGAAA TCCTGGAGAA ATCCTGAAAG
 601 GAGTTCTGCA TGGAACGAGA TTCAACATAG GAGACAANAT CCGGTACAGC
 651 TGCCTCCCTG GCTACATCTT GGAAGGCCAC GCCATCCTGA CCTGCATCGT
 701 CAGCCCAGGA AATGGTGCAT CGTGGGACTT CCCAGCTCCC TTTTGCAGAG
 751 CTGAGGGAGC CTGCGGAGGA ACCTTACGCG GGACCAGCAG CTCCATCTCC
 801 AGCCCGCACT TCCCTTCAGA GTACGAGAAC AACGCGGACT GCACCTGGAC
 851 CATTCTGGCT GAGCCCGGGG ACACCATTGC GCTGGTCTTC ACTGACTTTC
 901 AGCTAGAAGA AGGATATGAT TTCTTAGAGA TCAGTGGCAC GGAAGCTCCA
 951 TCCATATGGC TAACTGGCAT GAACCTCCCC TCTCCAGTTA TCAGTAGCAA
1001 GAATTGGCTA CGACTCCATT TCACCTCTGA CAGCAACCAC CGACGCAAAG
1051 GATTTAACGC TCAGTTCCAA GTGAAAAAGG CGATTGAGTT GAAGTCAAGA
1101 GGAGTCAAGA TGCTGCCCAG CAAGGATGGA AGCCATAAAA ACTCTGTCTT
1151 GAGCCAAGGA GGTGTTGCAT TGGTCTCTGA CATGTGTCCA GATCCTGGGA
1201 TTCCAGAAAA TGGTAGAAGA GCAGGTTCCG ACTTCAGGGT TGGTGCAAAT
1251 GTACAGTTTT CATGTGAGGA CAATTACGTG CTCCAGGGAT CTAAAAGCAT
1301 CACCTGTCAG AGAGTTACAG AGACGCTCGC TGCTTGGAGT GACCACAGGC
1351 CCATCTGCCG AGCGAGAACA TGTGGATCCA ATCTGCGTGG GCCCAGCGGC
1401 GTCATTACCT CCCCTAATTA TCCGGTTCAG TATGAAGATA ATGCACACTG
1451 TGTGTGGGTC ATCACCACCA CCGACCCGGA CAAGGTCATC AAGCTTGCCT
1501 TNGAAGAGTT TGAGCTGGAG CGAGGCTATG ACACCCTNAC GGTTGGTGAT
1551 GCTGGGAAGG TGGGAGACAC CAGATCGGTC TTGTANGTGC TCACGGGATC
1601 CAGTGTTCCT GACCTCATTG TGAGCATGAG CAACCAGATG TGGCTACATC
1651 TGCAGTCGGA TGATAGCATT GGCTCACCTG GGTTTAAAGC TGTTTACCAA
1701 GAAATTGAAA AGGGAGGGTG TGGGGATCCT GGAATCCCCG CCTATGGGAA
1751 GCGGACGGGC AGCAGTTTCC TCCATGGAGA TNCACTNACC TTTGAATGCC
1801 CGGCGGCCTT TGAGCTGGTG GGGGAGAGAG TTATCACCTG TCAGCAGAAC
1851 AATCAGTGGT CTGGCAACAA GCCCAGCTGT GTATTTTCAT GTTTCTTCAA
1901 CTTTACGGCA TCATCTGGGA TTATTCTGTC ACCAAATTAT CCAGAGGAAT
1951 ATGGGAACAA CATGAACTGT GTCTGGTTGA TTATCTCGGA GCCAGGAAGT
2001 CGAATTCACC TAATCTTTAA TGATTTTGAT GTTGAGCCTC AATTTGACTT
2051 TCTCGCGGTC AAGGATGATG GCATTTCTGA CATAACTGTC CTGGGTACTT
2101 TTTCTGGCAA TGAAGTGCCT TCCCAGCTGG CCAGCAGTGG GCATATAGTT
2151 CGCTTGGAAT TTCAGTCTGA CCATTCCACT ACTGGCAGAG GGTTNAACAT
2201 CACTTACACC ACNTTTGGTC AGAATGAGTG CCATGATCCT GGCATTCCTA
2251 TAAACGGACG ACGTTTTGGT GACAGGTTTC TACTCGGGAG CTCGGTTTCT
2301 TTCCACTGTG ATGATGGCTT TGTCAAGACC CAGGGATCCG AGTCCATTAC
2351 CTGCATACTG CAAGACGGGA ACGTGGTCTG GAGCTCCACC GTGCCCCGCT
2401 GTGAAGCTCC ATGTGGTGGA CATCTGACAG CGTCCAGCGG AGTCATTTTG
2451 CCTCCTGGAT GGCCAGGATA TTATAAGGAT TCTTTACATT GTGAATGGAT
2501 AATTGAAGCA AAACCAGGCC ACTCTATCAA AATAACTTTT GACAGATTTC
2551 AGACAGAGGT CAATTATGAC ACCTTGGAGG TCAGAGATGG GCCAGCCAGT
2601 TCGTCCCCAC TGATCGGCGA GTACCACGGC ACCCAGGCAC CCCAGTTCCT
2651 CATCAGCACC GGGAACTTCA TGTACCTGCT ATTCACCACT GACAACAGCC
2701 GCTCCAGCAT CGGCTTCCTC ATCCACTATG AGAGTGTGAC GCTTGAGTCG
2751 GATTCCTGCC TGGACCCGGG CATCCCTGTG AACGGCCATC GCCACGGTGG
2801 AGACTTTGGC ATCAGGTCCA CAGTGACTTT CAGCTGTGAC CCGGGGTACA
2851 CACTAAGTGA CGACGAGCCC CTCGTCTGTG AGAGGAACCA CCAGTGGAAC
2901 CACGCCTTGC CCAGCTGCGA CGCTCTATGT GGAGGCTACA TCCAAGGGAA
2951 GAGTGGAACA GTCCTTTCTC CTGGGTTTCC AGATTTTTAT CCAAACTCTC
```

3001	TAAACTGCAC	GTGGACCATT	GAAGTGTCTC	ATGGGAAAGG	AGTTCAAATG
3051	ATCTTTCACA	CCTTTCATCT	TGAGAGTTCC	CACGACTATT	TACTGATCAC
3101	AGAGGATGGA	AGTTTTTCCG	AGCCCGTTGC	CAGGCTCACC	GGGTCGGTGT
3151	TGCCTCATAC	GATCAAGGCA	GGCCTGTTNG	GAAACTTCAC	TGCCCAGCTT
3201	CGGTTTATAT	CAGACTTCTC	AATTTCGTAC	GAGGGCTTCA	ATATCACATT
3251	TTCAGAATAT	GACCTGGAGC	CATGTGATGA	TCCTGGAGTC	CCTGCCTTCA
3301	GCCGAAGAAT	TGGTTTTCAC	TTTGGTGTGG	GAGACTCTCT	GACGTTTTCC
3351	TGCTTCCTGG	GATATCGTTT	AGAAGGTGCC	ACCAAGCTTA	CCTGCCTGGG
3401	TGGGGGCCGC	CGTGTGTGGA	GTGCACCTCT	GCCAAGGTGT	GTGGCCGAAT
3451	GTGGAGCAAG	TGTCAAAGGA	AATGAAGGAA	CATTACTGTC	TCCAAATTTT
3501	CCATCCAATT	ATGATAATAA	CCATGAGTGT	ATCTATAAAA	TAGAAACAGA
3551	AGCCGGCAAG	GGCATCCACC	TTAGAACACG	AAGCTTCCAG	CTGTTTGAAG
3601	GAGATACTCT	AAAGGTATAT	GATGGAAAAG	ACAGTTCCTC	ACGTCCACTG
3651	GGCACGTTCA	CTAAAAAATGA	ACTTCTGGGG	CTGATCCTAA	ACAGCACATC
3701	CAATCACCTG	TGGCTAGAGT	TCAACACCAA	TGGATCTGAC	ACCGACCAAG
3751	GTTTTCAACT	CACCTATACC	AGTTTTGATC	TGGTAAAATG	TGAGGATCCG
3801	GGCATCCCTA	ACTACGGCTA	TAGGATCCGT	GATGAAGGCC	ACTTTACCGA
3851	CACTGTAGTT	CTGTACAGTT	GCAACCCGGG	GTACGCCATG	CATGGCAGCA
3901	ACACCCTGAC	CTGTTTGAGT	GGAGACAGGA	GAGTGTGGGA	CAAACCACTA
3951	CCTTCGTGCA	TAGCGGAATG	TGGTGGTCAG	ATCCATGCAG	CCACATCAGG
4001	ACGAATATTG	TCCCCTGGCT	ATCCAGCTCC	GTATGACAAC	AACCTCCACT
4051	GCACCTGGAT	TATAGAGGCA	GACCCAGGAA	AGACCATTAG	CCTCCATTTC
4101	ATTGTTTTCG	ACACGGAGAT	GGCTCACGAC	ATCCTCAAGG	TCTGGGACGG
4151	GCCGGTGGAC	AGTGACATCC	TGCTGAAGGA	GTGGAGTGGC	TCCGCCCTTC
4201	CGGAGGACAT	CCACAGCACC	TTCAACTCAC	TCACCCTGCA	GTTCGACAGC
4251	GACTTCTTCA	TCAGCAAGTC	TGGCTTCTCC	ATCCAGTTCT	CCACCTCAAT
4301	TGCAGCCACC	TGTAACGATC	CAGGTATGCC	CCAAAATGGC	ACCCGCTATG
4351	GAGACAGCAG	AGAGGCTGGA	GACACCGTCA	CATTCCAGTG	TGACCCTGGC
4401	TATCAGCTCC	AAGGACAAGC	CAAAATCACC	TGTGTGCAGC	TGAATAACCG
4451	GTTCTTTTGG	CAACCAGACC	CTCCTACATG	CATAGCTGCT	TGTGGAGGGA
4501	ATCTGACGGG	CCCAGCAGGT	GTTATTTTGT	CACCCAACTA	CCCACAGCCG
4551	TATCCTCCTG	GGAAGGAATG	TGACTGGAGA	GTAAAAGTGA	ACCCGGACTT
4601	TGTCATCGCC	TTGATATTCA	AAAGTTTCAA	CATGGAGCCC	AGCTATGACT
4651	TCCTACACAT	CTATGAAGGG	GAAGATTCCA	ACAGCCCCCT	CATTGGGAGT
4701	TACCAGGGCT	CTCAGGCCCC	AGAAAGAATA	GAGAGTAGCG	GAAACAGCCT
4751	GTTTCTGGCA	TTTCGGAGTG	ATGCCTCCGT	GGGCCTTTCA	GGGTTCGCCA
4801	TTGAATTTAA	AGAGAAACCA	CGGGAAGCTT	GTTTTGACCC	AGGAAATATA
4851	ATGAATGGGA	CAAGAGTTGG	AACAGACTTC	AAGCTTGGCT	CCACCATCAC
4901	CTACCAGTGT	GACTCTGGCT	ATAAGATTCT	TGACCCCTCA	TCCATCACCT
4951	GTGTGATTGG	GGCTGATGGG	AAACCCTCCT	GGGACCAAGT	GCTGCCCTCC
5001	TGCAATGCTC	CCTGTGGAGG	CCAGTACACG	GGATCAGAAG	GGGTAGTTTT
5051	ATCACCAAAC	TACCCCCATA	ATTACACAGC	TGGTCAAATA	TGCCTCTATT
5101	CCATCACGGT	ACCAAAGGAA	TTCGTGGTCT	TTGGACAGTT	TGCCTATTTC
5151	CAGACAGCCC	TGAATGATTT	GGCAGAATTA	TTTGATGGAA	CCCATGCACA
5201	GGCCAGACTT	CTCAGCTCAC	TCTCGGGGTC	TCACTCAGGG	GAAACATTGC
5251	CCTTGGCTAC	GTCAAATCAA	ATTCTGCTCC	GATTCAGTGC	AAAGAGCGGT
5301	GCCTCTGCCC	GCGGCTTCCA	CTTCGTGTAT	CAAGCTGTTC	CTCGTACCAG
5351	TGACACCCAA	TGCAGCTCTG	TCCCCGAGCC	CAGATACGGA	AGGAGAATTG
5401	GTTCTGAGTT	TTCTGCCGGC	TCCATCGTCC	GATTCGAGTG	CAACCCGGGA
5451	TACCTGCTTC	AGGGTTCCAC	GGCGCTCCAC	TGCCAGTCCG	TGCCCAACGC
5501	CTTGGCACAG	TGGAACGACA	CGATCCCCAG	CTGTGTGGTA	CCCTGCAGTG
5551	GCAATTTCAC	TCAACGAAGA	GGTACAATCC	TGTCCCCCGG	CTACCCTGAG
5601	CCATACGGAA	ACAACTTGAA	CTGTATATGG	AAGATCATAG	TTACGGAGGG
5651	CTCGGGAATT	CAGATCCAAG	TGATCAGTTT	TGCCACGGAG	CAGAACTGGG
5701	ACTCCCTTGA	GATCCACGAT	GGTGGGGATG	TGACCGCACC	CAGACTGGGA
5751	AGCTTCTCAG	GCACCACAGT	ACCGGCACTG	CTGAACAGTA	CTTCCAACCA
5801	ACTCTACCTG	CATTTCCAGT	CTGACATTAG	TGTGGCAGCT	GCTGGTTTCC
5851	ACCTGGAATA	CAAAACTGTA	GGTCTTGCTG	CATGCCAAGA	ACCAGCCCTC
5901	CCCAGCAACA	GCATCAAAAT	CGGAGATCGG	TACATGGTGA	ACGACGTGCT
5951	CTCCTTCCAG	TGCGAGCCCG	GGTACACCCT	GCAGGGCCGT	TCCCACATTT
6001	CCTGTATGCC	AGGGACCGTT	CGCCGTTGGA	ACTATCCGTC	TCCCCTGTGC
6051	ATTGCAACCT	GTGGAGGGAC	GCTGAGCACC	TTGGGTGGTG	TGATCCTGAG
6101	CCCCGGCTTC	CCAGGTTCTT	ACCCCAACAA	CTTAGACTGC	ACCTGGAGGA

```
6151  TCTCATTACC CATCGGCTAT GGTGCACATA TTCAGTTTCT GAATTTTTCT
6201  ACCGAAGCTA ATCATGACTT CCTTGAAATT CAAAATGGAC CTTACCACAC
6251  CAGCCCCATG ATTGGACAAT TTAGCGGCAC GGATCTCCCC GCGGCCCTGC
6301  TGAGCACAAC GCATGAAACC CTCATCCACT TTTATAGTGA CCATTCGCAA
6351  AACCGGCAAG GATTTAAACT TGCTTACCAA GCCTATGAAT TACAGAACTG
6401  TCCAGATCCA CCCCCATTTC AGAATGGGTA CATGATCAAC TCGGATTACA
6451  GCGTGGGGCA ATCAGTATCT TTCGAGTGTT ATCCTGGGTA CATTCTAATA
6501  GGCCATCCTG TCCTCACTTG TCAGCATGGG ATCAACAGAA ACTGGAACTA
6551  CCCTTTTCCA AGATGTGATG CCCCTTGTGG GTACAACGTA ACTTCTCAGA
6601  ACGGCACCAT CTACTCCCCT GGCTTTCCTG ATGAGTATCC GATCCTGAAG
6651  GACTGCATTT GGCTCATCAC GGTGCCTCCA GGGCACGGAG TTTACATCAA
6701  CTTCACCCTG TTACAGACGG AAGCTGTCAA CGATTACATT GCTGTTTGGG
6751  ACGGTCCCGA TCAGAACTCA CCCCAGCTGG GAGTTTTCAG TGGCAACACA
6801  GCCCTCGAAA CGGCGTATAG CTCCACCAAC CAAGTCCTGC TCAAGTTCCA
6851  CAGCGACTTT TCAAATGGAG GCTTCTTTGT CCTCAATTTC CACGGTCAGT
6901  TGATTTTCAC TCCGTTAGTT AAGACTGAGA ATTCCATGTG GTGTTTACTG
6951  CAGTGTTGTC CCACGCCTTG TTTCCAGCTG AAGTTTCTTG ATTCAGCCGA
7001  GGGCGTGTAT GATTCTTTTG CACTGGAGGC CAGCGTTTCC TGTGGTCCTT
7051  TTTTTGTTTA ATGATGTCTT TATTATTTCA CATCGTATCC AGCTTGGATT
7101  TATTCCAAGA TACATGTATC CTAAGTGAAA CTCTAAGATG AAGACCATTG
7151  AAAGAGATTT GGTACCTTTT ATAGATTTAC TCATCCCTGT CTCAAGATAA
7201  GGTGTTATAG CAAATGTCAT GTAACTATAA ATGGTGTGAA AGCAAACCTC
7251  CAATAATCCT GGGAATGCAC TCTAAACGAT ATGTAGAACA TCTGTCAATC
7301  NATCGCTTAT CTCTCACGAA CAC
```

SEQ ID NO:6

5R23V2

AGCTTGTGCCCTTTCCACCTGCATTTCTGATCTAAGTTAGGTAGGGGGCTGCTCTCTGGTC
AGCAAGGAAGGGAGATCAAAGGATGGAGGCGGGACTCTGCCCCTGCAGAAACCCTCCAG
TTTGCTGGAGTTGCCGGATTACATTGTTCCTCCCCGGTGTGCGGCGTGAGCTTCCCCCACC
CGAGCGCCCAACAAGTCTCCTTTCTCCAGCCTGCGCGCTGCTGCGCTGAGGCCGAATGAA
GCGCAGCACGGTGCGGGCAGCCCGAGGCCCCGAGGCTGGGCTCTGTCTGTCTGGGACTGC
GCCGTGCCCAGCCTCGGTCCCCTCTCTGTGGGTAAGGATGGTTGAGTCCAGCCTCCACGG
CAGCGGCTCCTTGTGCCACTAGCAGCCCTTCTTCTGCGCTCTCCGCCTTTTCTCTCTAGAC
TGGATCTCTCCTCCCCCCGCGCCCCCCTCCCCGCATCTCCCACTCGCTGGCTCTCTCTCCA
GCTGCCTCCTCTCCAGGTCTCTCCTGGCTGCGCGCGCTCCTCTCCCCGCTTCTCCCCCTCCC
GCAGCCTCGCCGCCTTGGTGCCTTCCTGCCCGGCTCGGCCGGCGCTCGTCCCCGGCCCCG
GCCCCGCCAGCCCGGGTCTCCGCGCTCGGAGCAGCTCAGCCCTGCAGTGGCTCGGGACCC
GATGCTATGAGAGGGAAGCGAGCCGGGCGCCCAGACCTTCAGGAGGCGTCGGATGCGCG
GCGGGTCTTGGGACCGGGCTCTCTCTCCGGCTCGCCTTGCCCTCGGGTGATTATTTGGCTC
CGCTCATAGCCCTGCCTTCCTCGGAGGAGCCATCGGTGTCGCGTGCGTGTGGAGTATCTG
CAGACATGACTGCGTGGAGGAGATTCCAGTCGCTGCTCCTGCTTCTCGGGCTGCTGGTGC
TGTGCGCGAGGCTCCTCACTGCAGCGAAGGGTCAGAACTGTGGAGGCTTAGTCCAGGGTC
CCAATGGCACTATTGAGAGCCCAGGGTTTCCTCACGGGTATCCGAACTATGCCAACTGCA
CCTGGATCATCATCACGGGCGAGCGCAATAGGATACAGTTGTCCTTCCATACCTTTGCTCT
TGAAGAAGATTTTGATATTTTATCAGTTTACGATGGACAGCCTCAACAAGGGAATTTAAA
AGTGAGATTATCGGGATTTCAGCTGCCCTCCTCTATAGTGAGTACAGGATCTATCCTCACT
CTGTGGTTCACGACAGACTTCGCTGTGAGTGCCCAAGGTTTCAAAGCATTATATGAAGTT
TTACCTAGCCACACTTGTGGAAATCCTGGAGAAATCCTGAAAGGAGTTCTGCATGGAACG
AGATTCAACATAGGAGACAANATCCGGTACAGCTGCCTCCCTGGCTACATCTTGGAAGGC
CACGCCATCCTGACCTGCATCGTCAGCCCAGGAAATGGTGCATCGTGGGACTTCCCAGCT
CCCTTTTGCAGAGCTGAGGGAGCCTGCGGAGGAACCTTACGCGGGACCAGCAGCTCCATC
TCCAGCCCGCACTTCCCTTCAGAGTACGAGAACAACGCGGACTGCACCTGGACCATTCTG
GCTGAGCCCGGGGACACCATTGCGCTGGTCTTCACTGACTTTCAGCTAGAAGAAGGATAT
GATTTCTTAGAGATCAGTGGCACGGAAGCTCCATCCATATGGCTAACTGGCATGAACCTC
CCCTCTCCAGTTATCAGTAGCAAGAATTGGCTACGACTCCATTTCACCTCTGACAGCAACC
ACCGACGCAAAGGATTTAACGCTCAGTTCCAAGTGAAAAAGGCGATTGAGTTGAAGTCA
AGAGGAGTCAAGATGCTGCCCAGCAAGGATGGAAGCCATAAAAACTCTGTCTTGAGCCA
AGGAGGTGTTGCATTGGTCTCTGACATGTGTCCAGATCCTGGGATTCCAGAAAATGGTAG
AAGAGCAGGTTCCGACTTCAGGGTTGGTGCAAATGTACAGTTTTCATGTGAGGACAATTA
CGTGCTCCAGGGATCTAAAAGCATCACCTGTCAGAGAGTTACAGAGACGCTCGCTGCTTG
GAGTGACCACAGGCCCATCTGCCGAGCGAGAACATGTGGATCCAATCTGCGTGGGCCCAG
CGGCGTCATTACCTCCCCTAATTATCCGGTTCAGTATGAAGATAATGCACACTGTGTGTG
GGTCATCACCACCACCGACCCGGACAAGGTCATCAAGCTTGCCTTNGAAGAGTTTGAGCT
GGAGCGAGGCTATGACACCCTNACGGTTGGTGATGCTGGGAAGGTGGGAGACACCAGAT
CGGTCTTGTANGTGCTCACGGGATCCAGTGTTCCTGACCTCATTGTGAGCATGAGCAACC
AGATGTGGCTACATCTGCAGTCGGATGATAGCATTGGCTCACCTGGGTTTAAAGCTGTTT
ACCAAGAAATTGAAAAGGGAGGGTGTGGGGATCCTGGAATCCCCGCCTATGGGAAGCGG
ACGGGCAGCAGTTTCCTCCATGGAGATNCACTNACCTTTGAATGCCCGGCGGCCTTTGAG
CTGGTGGGGGAGAGAGTTATCACCTGTCAGCAGAACAATCAGTGGTCTGGCAACAAGCCC
AGCTGTGTATTTTCATGTTTCTTCAACTTTACGGCATCATCTGGGATTATTCTGTCACCAA
ATTATCCAGAGGAATATGGGAACAACATGAACTGTGTCTGGTTGATTATCTCGGAGCCAG
GAAGTCGAATTCACCTAATCTTTAATGATTTTGATGTTGAGCCTCAATTTGACTTTCTCGC
GGTCAAGGATGATGGCATTTCTGACATAACTGTCCTGGGTACTTTTTCTGGCAATGAAGT
GCCTTCCCAGCTGGCCAGCAGTGGGCATATAGTTCGCTTGGAATTTCAGTCTGACCATTCC
ACTACTGGCAGAGGGTTNAACATCACTTACACCACNTTTGGTCAGAATGAGTGCCATGAT
CCTGGCATTCCTATAAACGGACGACGTTTTGGTGACAGGTTTCTACTCGGGAGCTCGGTT
TCTTTCCACTGTGATGATGGCTTTGTCAAGACCCAGGGATCCGAGTCCATTACCTGCATAC
TGCAAGACGGGAACGTGGTCTGGAGCTCCACCGTGCCCCGCTGTGAAGCTCCATGTGGTG
GACATCTGACAGCGTCCAGCGGAGTCATTTTGCCTCCTGGATGGCCAGGATATTATAAGG
ATTCTTTACATTGTGAATGGATAATTGAAGCAAAACCAGGCCACTCTATCAAAATAACTT

TTGACAGATTTCAGACAGAGGTCAATTATGACACCTTGGAGGTCAGAGATGGGCCAGCCA
GTTCGTCCCCACTGATCGGCGAGTACCACGGCACCCAGGCACCCCAGTTCCTCATCAGCA
CCGGGAACTTCATGTACCTGCTATTCACCACTGACAACAGCCGCTCCAGCATCGGCTTCCT
CATCCACTATGAGAGTGTGACGCTTGAGTCGGATTCCTGCCTGGACCCGGGCATCCCTGT
GAACGGCCATCGCCACGGTGGAGACTTTGGCATCAGGTCCACAGTGACTTTCAGCTGTGA
CCCGGGGTACACACTAAGTGACGACGAGCCCCTCGTCTGTGAGAGGAACCACCAGTGGA
ACCACGCCTTGCCCAGCTGCGACGCTCTATGTGGAGGCTACATCCAAGGGAAGAGTGGAA
CAGTCCTTTCTCCTGGGTTTCCAGATTTTTATCCAAACTCTCTAAACTGCACGTGGACCAT
TGAAGTGTCTCATGGGAAAGGAGTTCAAATGATCTTTCACACCTTTCATCTTGAGAGTTCC
CACGACTATTTACTGATCACAGAGGATGGAAGTTTTTCCGAGCCCGTTGCCAGGCTCACC
GGGTCGGTGTTGCCTCATACGATCAAGGCAGGCCTGTTNGGAAACTTCACTGCCCAGCTT
CGGTTTATATCAGACTTCTCAATTTCGTACGAGGGCTTCAATATCACATTTTCAGAATATG
ACCTGGAGCCATGTGATGATCCTGGAGTCCCTGCCTTCAGCCGAAGAATTGGTTTTCACTT
TGGTGTGGGAGACTCTCTGACGTTTTCCTGCTTCCTGGGATATCGTTTAGAAGGTGCCACC
AAGCTTACCTGCCTGGGTGGGGGCCGCCGTGTGTGGAGTGCACCTCTGCCAAGGTGTGTG
GCCGAATGTGGAGCAAGTGTCAAAGGAAATGAAGGAACATTACTGTCTCCAAATTTTCCA
TCCAATTATGATAATAACCATGAGTGTATCTATAAAATAGAAACAGAAGCCGGCAAGGGC
ATCCACCTTAGAACACGAAGCTTCCAGCTGTTTGAAGGAGATACTCTAAAGGTATATGAT
GGAAAAGACAGTTCCTCACGTCCACTGGGCACGTTCACTAAAAATGAACTTCTGGGGCTG
ATCCTAAACAGCACATCCAATCACCTGTGGCTAGAGTTCAACACCAATGGATCTGACACC
GACCAAGGTTTTCAACTCACCTATACCAGTTTTGATCTGGTAAAATGTGAGGATCCGGGC
ATCCCTAACTACGGCTATAGGATCCGTGATGAAGGCCACTTTACCGACACTGTAGTTCTG
TACAGTTGCAACCCGGGGTACGCCATGCATGGCAGCAACACCCTGACCTGTTTGAGTGGA
GACAGGAGAGTGTGGGACAAACCACTACCTTCGTGCATAGCGGAATGTGGTGGTCAGAT
CCATGCAGCCACATCAGGACGAATATTGTCCCCTGGCTATCCAGCTCCGTATGACAACAA
CCTCCACTGCACCTGGATTATAGAGGCAGACCCAGGAAAGACCATTAGCCTCCATTTCAT
TGTTTTCGACACGGAGATGGCTCACGACATCCTCAAGGTCTGGGACGGGCCGGTGGACAG
TGACATCCTGCTGAAGGAGTGGAGTGGCTCCGCCCTTCCGGAGGACATCCACAGCACCTT
CAACTCACTCACCCTGCAGTTCGACAGCGACTTCTTCATCAGCAAGTCTGGCTTCTCCATC
CAGTTCTCCACCTCAATTGCAGCCACCTGTAACGATCCAGGTATGCCCCAAAATGGCACC
CGCTATGGAGACAGCAGAGAGGCTGGAGACACCGTCACATTCCAGTGTGACCCTGGCTAT
CAGCTCCAAGGACAAGCCAAAATCACCTGTGTGCAGCTGAATAACCGGTTCTTTTGGCAA
CCAGACCCTCCTACATGCATAGCTGCTTGTGGAGGGAATCTGACGGGCCCAGCAGGTGTT
ATTTTGTCACCCAACTACCCACAGCCGTATCCTCCTGGGAAGGAATGTGACTGGAGAGTA
AAAGTGAACCCGGACTTTGTCATCGCCTTGATATTCAAAAGTTTCAACATGGAGCCCAGC
TATGACTTCCTACACATCTATGAAGGGGAAGATTCCAACAGCCCCCTCATTGGGAGTTAC
CAGGGCTCTCAGGCCCCAGAAAGAATAGAGAGTAGCGGAAACAGCCTGTTTCTGGCATTT
CGGAGTGATGCCTCCGTGGGCCTTTCAGGGTTCGCCATTGAATTTAAAGAGAAACCACGG
GAAGCTTGTTTTGACCCAGGAAATATAATGAATGGGACAAGAGTTGGAACAGACTTCAAG
CTTGGCTCCACCATCACCTACCAGTGTGACTCTGGCTATAAGATTCTTGACCCCTCATCCA
TCACCTGTGTGATTGGGGCTGATGGGAAACCCTCCTGGGACCAAGTGCTGCCCTCCTGCA
ATGCTCCCTGTGGAGGCCAGTACACGGGATCAGAAGGGGTAGTTTTATCACCAAACTACC
CCCATAATTACACAGCTGGTCAAATATGCCTCTATTCCATCACGGTACCAAAGGAATTCG
TGGTCTTTGGACAGTTTGCCTATTTCCAGACAGCCCTGAATGATTTGGCAGAATTATTTGA
TGGAACCCATGCACAGGCCAGACTTCTCAGCTCACTCTCGGGGTCTCACTCAGGGGAAAC
ATTGCCCTTGGCTACGTCAAATCAAATTCTGCTCCGATTCAGTGCAAAGAGCGGTGCCTCT
GCCCGCGGCTTCCACTTCGTGTATCAAGCTGTTCCTCGTACCAGTGACACCCAATGCAGCT
CTGTCCCCGAGCCCAGATACGGAAGGAGAATTGGTTCTGAGTTTTCTGCCGGCTCCATCG
TCCGATTCGAGTGCAACCCGGGATACCTGCTTCAGGGTTCCACGGCGCTCCACTGCCAGT
CCGTGCCCAACGCCTTGGCACAGTGGAACGACACGATCCCCAGCTGTGTGGTACCCTGCA
GTGGCAATTTCACTCAACGAAGAGGTACAATCCTGTCCCCCGGCTACCCTGAGCCATACG
GAAACAACTTGAACTGTATATGGAAGATCATAGTTACGGAGGGCTCGGGAATTCAGATCC
AAGTGATCAGTTTTGCCACGGAGCAGAACTGGGACTCCCTTGAGATCCACGATGGTGGGG
ATGTGACCGCACCCAGACTGGGAAGCTTCTCAGGCACCACAGTACCGGCACTGCTGAACA
GTACTTCCAACCAACTCTACCTGCATTTCCAGTCTGACATTAGTGTGGCAGCTGCTGGTTT
CCACCTGGAATACAAAACTGTAGGTCTTGCTGCATGCCAAGAACCAGCCCTCCCCAGCAA
CAGCATCAAAATCGGAGATCGGTACATGGTGAACGACGTGCTCTCCTTCCAGTGCGAGCC

CGGGTACACCCTGCAGGGCCGTTCCCACATTTCCTGTATGCCAGGGACCGTTCGCCGTTG
GAACTATCCGTCTCCCCTGTGCATTGCAACCTGTGGAGGGACGCTGAGCACCTTGGGTGG
TGTGATCCTGAGCCCCGGCTTCCCAGGTTCTTACCCCAACAACTTAGACTGCACCTGGAG
GATCTCATTACCCATCGGCTATGGTGCACATATTCAGTTTCTGAATTTTTCTACCGAAGCT
AATCATGACTTCCTTGAAATTCAAAATGGACCTTACCACACCAGCCCCATGATTGGACAA
TTTAGCGGCACGGATCTCCCCGCGGCCCTGCTGAGCACAACGCATGAAACCCTCATCCAC
TTTTATAGTGACCATTCGCAAAACCGGCAAGGATTTAAACTTGCTTACCAAGCCTATGAA
TTACAGAACTGTCCAGATCCACCCCCATTTCAGAATGGGTACATGATCAACTCGGATTAC
AGCGTGGGGCAATCAGTATCTTTCGAGTGTTATCCTGGGTACATTCTAATAGGCCATCCT
GTCCTCACTTGTCAGCATGGGATCAACAGAAACTGGAACTACCCTTTTCCAAGATGTGAT
GCCCCTTGTGGGTACAACGTAACTTCTCAGAACGGCACCATCTACTCCCCTGGCTTTCCTG
ATGAGTATCCGATCCTGAAGGACTGCATTTGGCTCATCACGGTGCCTCCAGGGCACGGAG
TTTACATCAACTTCACCCTGTTACAGACGGAAGCTGTCAACGATTACATTGCTGTTTGGGA
CGGTCCCGATCAGAACTCACCCCAGCTGGGAGTTTTCAGTGGCAACACAGCCCTCGAAAC
GGCGTATAGCTCCACCAACCAAGTCCTGCTCAAGTTCCACAGCGACTTTTCAAATGGAGG
CTTCTTTGTCCTCAATTTCCACGGTCAGTTGATTTTCACTCCGTTAGTTAAGACTGAGAAT
TCCATGTGGTGTTTACTGCAGTGTTGTCCCACGCCTTGTTTCCAGCTGAAGTTTCTTGATT
CAGCCGAGGGCGTGTATGATTCTTTTGCACTGGAGGCCAGCGTTTCCTGTGGTCCTTTTTT
TGTTTAATGATGTCTTTATTATTTCACATCGTATCCAGCTTGGATTTATTCCAAGATACAT
GTATCCTAAGTGAAACTCTAAGATGAAGACCATTGAAAGAGATTTGGTACCTTTTATAGA
TTTACTCATCCCTGTCTCAAGATAAGGTGTTATAGCAAATGTCATGTAACTATAAATGGTG
TGAAAGCAAACCTCCAATAATCCTGGGAATGCACTCTAAACGATATGTAGAACATCTGTC
AATCNATCGCTTATCTCTCACGAACACN

SEQ ID NO:7

5R2_OC147

AGCTTGTGCCCTTTCCACCTGCATTTCTGATCTAAGTTAGGTAGGGGGCTGCTCTCTGGTCAGCAAGG
AAGGGAGATCAAAGGATGGAGGCGGGACTCTGCCCCTGCAGAAACCCTCCAGTTTGCTGGAGTTGCCG
GATTACATTGTTCCTCCCCGGTGTGCGGCGTGAGCTTCCCCCACCCGAGCGCCCAACAAGTCTCCTTT
CTCCAGCCTGCGCGCTGCTGCGCTGAGGCCGAATGAAGCGCAGCACGGTGCGGGCAGCCCGAGGCCCC
GAGGCTGGGCTCTGTCTGTCTGGGACTGCGCCGTGCCCAGCCTCGGTCCCCTCTCTGTGGGTAAGGAT
GGTTGAGTCCAGCCTCCACGGCAGCGGCTCCTTGTGCCACTAGCAGCCCTTCTTCTGCGCTCTCCGCC
TTTTCTCTCTAGACTGGATCTCTCCTCCCCCCGCGCCCCCCTCCCCGCATCTCCCACTCGCTGGCTCT
CTCTCCAGCTGCCTCCTCTCCAGGTCTCTCCTGGCTGCGCGCGCTCCTCTCCCCGCTTCTCCCCCTCC
CGCAGCCTCGCCGCCTTGGTGCCTTCCTGCCCGGCTCGGCCGGCGCTCGTCCCCGGCCCCGGCCCCGC
CAGCCCGGGTCTCCGCGCTCGGAGCAGCTCAGCCCTGCAGTGGCTCGGGACCCGATGCTATGAGAGGG
AAGCGAGCCGGGCGCCCAGACCTTCAGGAGGCGTCGGATGCGCGGCGGGTCTTGGGACCGGGCTCTCT
CTCCGGCTCGCCTTGCCCTCGGGTGATTATTTGGCTCCGCTCATAGCCCTGCCTTCCTCGGAGGAGCC
ATCGGTGTCGCGTGCGTGTGGAGTATCTGCAGACATGACTGCGTGGAGGAGATTCCAGTCGCTGCTCC
TGCTTCTCGGGCTGCTGGTGCTGTGCGCGAGGCTCCTCACTGCAGCGAAGGGTCAGAACTGTGGAGGC
TTAGTCCAGGGTCCCAATGGCACTATTGAGAGCCCAGGGTTTCCTCACGGGTATCCGAACTATGCCAA
CTGCACCTGGATCATCATCACGGGCGAGCGCAATAGGATACAGTTGTCCTTCCATACCTTTGCTCTTG
AAGAAGATTTTGATATTTTATCAGTTTACGATGGACAGCCTCAACAAGGGAATTTAAAAGTGAGATTA
TCGGGATTTCAGCTGCCCTCCTCTATAGTGAGTACAGGATCTATCCTCACTCTGTGGTTCACGACAGA
CTTCGCTGTGAGTGCCCAAGGTTTCAAAGCATTATATGAAGTTTTACCTAGCCACACTTGTGGAAATC
CTGGAGAAATCCTGAAAGGAGTTCTGCATGGAACGAGATTCAACATAGGAGACAAAATCCGGTACAGC
TGCCTCCCTGGCTACATCTTGGAAGGCCACGCCATCCTGACCTGCATCGTCAGCCCAGGAAATGGTGC
ATCGTGGGACTTCCCAGCTCCCTTTTGCAGAGCTGAGGGAGCCTGCGGAGGAACCTTACGCGGGACCA
GCAGCTCCATCTCCAGCCCGCACTTCCCTTCAGAGTACGAGAACAACGCGGACTGCACCTGGACCATT
CTGGCTGAGCCCGGGGACACCATTGCGCTGGTCTTCACTGACTTTCAGCTAGAAGAAGGATATGATTT
CTTAGAGATCAGTGGCACGGAAGCTCCATCCATATGGCTAACTGGCATGAACCTCCCCTCTCCAGTTA
TCAGTAGCAAGAATTGGCTACGACTCCATTTCACCTCTGACAGCAACCACCGACGCAAAGGATTTAAC
GCTCAGTTCCAAGTGAAAAAGGCGATTGAGTTGAAGTCAAGAGGAGTCAAGATGCTGCCCAGCAAGGA
TGGAAGCCATAAAAACTCTGTCTGTGAGTCCCTTTCCTTTCTATCTGAGGATTGATACGCCCTTGTAA
GCAGAGGAGAGAATGGAGCAGTG

SEQ ID NO:8

5R2_AW

AGCTTGTGCCCTTTCCACCTGCATTTCTGATCTAAGTTAGGTAGGGGGCTGCTCTCTGGTCAGCAAGG
AAGGGAGATCAAAGGATGGAGGCGGGACTCTGCCCCTGCAGAAACCCTCCAGTTTGCTGGAGTTGCCG
GATTACATTGTTCCTCCCCGGTGTGCGGCGTGAGCTTCCCCCACCCGAGCGCCCAACAAGTCTCCTTT
CTCCAGCCTGCGCGCTGCTGCGCTGAGGCCGAATGAAGCGCAGCACGGTGCGGGCAGCCCGAGGCCCC
GAGGCTGGGCTCTGTCTGTCTGGGACTGCGCCGTGCCCAGCCTCGGTCCCCTCTCTGTGGGTAAGGAT
GGTTGAGTCCAGCCTCCACGGCAGCGGCTCCTTGTGCCACTAGCAGCCCTTCTTCTGCGCTCTCCGCC
TTTTCTCTCTAGACTGGATCTCTCCTCCCCCCGCGCCCCCCTCCCCGCATCTCCCACTCGCTGGCTCT
CTCTCCAGCTGCCTCCTCTCCAGGTCTCTCCTGGCTGCGCGCGCTCCTCTCCCCGCTTCTCCCCCTCC
CGCAGCCTCGCCGCCTTGGTGCCTTCCTGCCCGGCTCGGCCGGCGCTCGTCCCCGGCCCCGGCCCCGC
CAGCCCGGGTCTCCGCGCTCGGAGCAGCTCAGCCCTGCAGTGGCTCGGGACCCGATGCTATGAGAGGG
AAGCGAGCCGGGCGCCCAGACCTTCAGGAGGCGTCGGATGCGCGGCGGGTCTTGGGACCGGGCTCTCT
CTCCGGCTCGCCTTGCCCTCGGGTGATTATTTGGCTCCGCTCATAGCCCTGCCTTCCTCGGAGGAGCC
ATCGGTGTCGCGTGCGTGTGGAGTATCTGCAGACATGACTGCGTGGAGGAGATTCCAGTCGCTGCTCC
TGCTTCTCGGGCTGCTGGTGCTGTGCGCGAGGCTCCTCACTGCAGCGAAGGGTCAGAACTGTGGAGGC
TTAGTCCAGGGTCCCAATGGCACTATTGAGAGCCCAGGGTTTCCTCACGGGTATCCGAACTATGCCAA
CTGCACCTGGATCATCATCACGGGCGAGCGCAATAGGATACAGTTGTCCTTCCATACCTTTGCTCTTG
AAGAAGATTTTGATATTTTATCAGTTTACGATGGACAGCCTCAACAAGGGAATTTAAAAGTGAGATTA
TCGGGATTTCAGCTGCCCTCCTCTATAGTGAGTACAGGATCTATCCTCACTCTGTGGTTCACGACAGA
CTTCGCTGTGAGTGCCCAAGGTTTCAAAGCATTATATGAAGTTTTACCTAGCCACACTTGTGGAAATC
CTGGAGAAATCCTGAAAGGAGTTCTGCATGGAACGAGATTCAACATAGGAGACAAAATCCGGTACAGC
TGCCTCCCTGGCTACATCTTGGAAGGCCACGCCATCCTGACCTGCATCGTCAGCCCAGGAAATGGTGC
ATCGTGGGACTTCCCAGCTCCCTTTTGCAGAGCTGAGGGAGCCTGCGGAGGAACCTTACGCGGGACCA
GCAGCTCCATCTCCAGCCCGCACTTCCCTTCAGAGTACGAGAACAACGCGGACTGCACCTGGACCATT
CTGGCTGAGCCCGGGGACACCATTGCGCTGGTCTTCACTGACTTTCAGCTAGAAGAAGGATATGATTT
CTTAGAGATCAGTGGCACGGAAGCTCCATCCATATGGCTAACTGGCATGAACCTCCCCTCTCCAGTTA
TCAGTAGCAAGAATTGGCTACGACTCCATTTCACCTCTGACAGCAACCACCGACGCAAAGGATTTAAC
GCTCAGTTCCAAGTGAAAAAGGCGATTGAGTTGAAGTCAAGAGGAGTCAAGATGCTGCCCAGCAAGGA
TGGAAGCCATAAAAACTCTGTCTGGCATCAGCAAGAGTTCAGCAAGTGCAGGAAGAAAAAGAGAGAGA
TCATGACAAGGAATGGGAGAATTTCCCTGACAGCCTCAGGAAACTTGCAGTTTGATAATTAAACAGAT
CAAGGTCACTCAGATGAGCTGATGGGACATGCTGTGTACGGAGGAGCATTTGCAGTTACAACACTTTG
TAGCCATGCAGGATGGGGCAATTAATCCAGAACCATTATTTAATAAAAAGATGATTTTTTAAATGTGA
AA

SEQ ID NO:9

protein sequence
>ORF:121..5598 Frame +1

```
MEAIKTLSGIWNNINHVTSEEDTFIMYLGKPWLQVKIQVSQGGVALVSDMCPDPGIPENGRRAGSDFR
VGANVQFSCEDNYVLQGSKSITCQRVTETLAAWSDHRPICRARTCGSNLRGPSGVITSPNYPVQYEDN
AHCVWVITTTDPDKVIKLAFEEFELERGYDTLTVGDAGKVGDTRSVLYVLTGSSVPDLIVSMSNQMWL
HLQSDDSIGSPGFKAVYQEIEKGGCGDPGIPAYGKRTGSSFLHGDTLTFECPAAFELVGERVITCQQN
NQWSGNKPSCVFSCFFNFTASSGIILSPNYPEEYGNNMNCVWLIISEPGSRIHLIFNDFDVEPQFDFL
AVKDDGISDITVLGTFSGNEVPSQLASSGHIVRLEFQSDHSTTGRGFNITYTTFGQNECHDPGIPING
RRFGDRFLLGSSVSFHCDDGFVKTQGSESITCILQDGNVVWSSTVPRCEAPCGGHLTASSGVILPPGW
PGYYKDSLHCEWIIEAKPGHSIKITFDRFQTEVNYDTLEVRDGPASSSPLIGEYHGTQAPQFLISTGN
FMYLLFTTDNSRSSIGFLIHYESVTLESDSCLDPGIPVNGHRHGGDFGIRSTVTFSCDPGYTLSDDEP
LVCERNHQWNHALPSCDALCGGYIQGKSGTVLSPGFPDFYPNSLNCTWTIEVSHGKGVQMIFHTFHLE
SSHDYLLITEDGSFSEPVARLTGSVLPHTIKAGLFGNFTAQLRFISDFSISYEGFNITFSEYDLEPCD
DPGVPAFSRRIGFHFGVGDSLTFSCFLGYRLEGATKLTCLGGGRRVWSAPLPRCVAECGASVKGNEGT
LLSPNFPSNYDNNHECIYKIETEAGKGIHLRTRSFQLFEGDTLKVYDGKDSSSRPLGTFTKNELLGLI
LNSTSNHLWLEFNTNGSDTDQGFQLTYTSFDLVKCEDPGIPNYGYRIRDEGHFTDTVVLYSCNPGYAM
HGSNTLTCLSGDRRVWDKPLPSCIAECGGQIHAATSGRILSPGYPAPYDNNLHCTWIIEADPGKTISL
HFIVFDTEMAHDILKVWDGPVDSDILLKEWSGSALPEDIHSTFNSLTLQFDSDFFISKSGFSIQFSTS
IAATCNDPGMPQNGTRYGDSREAGDTVTFQCDPGYQLQGQAKITCVQLNNRFFWQPDPPTCIAACGGN
LTGPAGVILSPNYPQPYPPGKECDWRVKVNPDFVIALIFKSFNMEPSYDFLHIYEGEDSNSPLIGSYQ
GSQAPERIESSGNSLFLAFRSDASVGLSGFAIEFKEKPREACFDPGNIMNGTRVGTDFKLGSTITYQC
DSGYKILDPSSITCVIGADGKPSWDQVLPSCNAPCGGQYTGSEGVVLSPNYPHNYTAGQICLYSITVP
KEFVVFGQFAYFQTALNDLAELFDGTHAQARLLSSLSGSHSGETLPLATSNQILLRFSAKSGASARGF
HFVYQAVPRTSDTQCSSVPEPRYGRRIGSEFSAGSIVRFECNPGYLLQGSTALHCQSVPNALAQWNDT
IPSCVVPCSGNFTQRRGTILSPGYPEPYGNNLNCIWKIIVTEGSGIQIQVISFATEQNWDSLEIHDGG
DVTAPRLGSFSGTTVPALLNSTSNQLYLHFQSDISVAAAGFHLEYKTVGLAACQEPALPSNSIKIGDR
YMVNDVLSFQCEPGYTLQGRSHISCMPGTVRRWNYPSPLCIATCGGTLSTLGGVILSPGFPGSYPNNL
DCTWRISLPIGYGAHIQFLNFSTEANHDFLEIQNGPYHTSPMIGQFSGTDLPAALLSTTHETLIHFYS
DHSQNRQGFKLAYQAYELQNCPDPPPFQNGYMINSDYSVGQSVSFECYPGYILIGHPP
```

SEQ ID NO:10

```
5G-3V1 Protein sequence          1801 AA
   1 MEAIKTLSGI WNNINHVTSE EDTFIMYLGK PWLQVKIQVS QGGVALVSDM
  51 CPDPGIPENG RRAGSDFRVG ANVQFSCEDN YVLQGSKSIT CQRVTETLAA
 101 WSDHRPICRA RTCGSNLRGP SGVITSPNYP VQYEDNAHCV WVITTTDPDK
 151 VIKLAFEEFE LERGYDTLTV GDAGKVGDTR SVLYVLTGSS VPDLIVSMSN
 201 QMWLHLQSDD SIGSPGFKAV YQEIEKGGCG DPGIPAYGKR TGSSFLHGDT
 251 LTFECPAAFE LVGERVITCQ QNNQWSGNKP SCVFSCFFNF TASSGIILSP
 301 NYPEEYGNNM NCVWLIISEP GSRIHLIFND FDVEPQFDFL AVKDDGISDI
 351 TVLGTFSGNE VPSQLASSGH IVRLEFQSDH STTGRGFNIT YTTFGQNECH
 401 DPGIPINGRR FGDRFLLGSS VSFHCDDGFV KTQGSESITC ILQDGNVVWS
 451 STVPRCEAPC GGHLTASSGV ILPPGWPGYY KDSLHCEWII EAKPGHSIKI
 501 TFDRFQTEVN YDTLEVRDGP ASSSPLIGEY HGTQAPQFLI STGNFMYLLF
 551 TTDNSRSSIG FLIHYESVTL ESDSCLDPGI PVNGHRHGGD FGIRSTVTFS
 601 CDPGYTLSDD EPLVCERNHQ WNHALPSCDA LCGGYIQGKS GTVLSPGFPD
 651 FYPNSLNCTW TIEVSHGKGV QMIFHTFHLE SSHDYLLITE DGSFSEPVAR
 701 LTGSVLPHTI KAGLFGNFTA QLRFISDFSI SYEGFNITFS EYDLEPCDDP
 751 GVPAFSRRIG FHFGVGDSLT FSCFLGYRLE GATKLTCLGG GRRVWSAPLP
 801 RCVAECGASV KGNEGTLLSP NFPSNYDNNH ECIYKIETEA GKGIHLRTRS
 851 FQLFEGDTLK VYDGKDSSSR PLGTFTKNEL LGLILNSTSN HLWLEFNTNG
 901 SDTDQGFQLT YTSFDLVKCE DPGIPNYGYR IRDEGHFTDT VVLYSCNPGY
 951 AMHGSNTLTC LSGDRRVWDK PLPSCIAECG GQIHAATSGR ILSPGYPAPY
1001 DNNLHCTWII EADPGKTISL HFIVFDTEMA HDILKVWDGP VDSDILLKEW
1051 SGSALPEDIH STFNSLTLQF DSDFFISKSG FSIQFSTSIA ATCNDPGMPQ
1101 NGTRYGDSRE AGDTVTFQCD PGYQLQGQAK ITCVQLNNRF FWQPDPPTCI
1151 AACGGNLTGP AGVILSPNYP QPYPPGKECD WRVKVNPDFV IALIFKSFNM
1201 EPSYDFLHIY EGEDSNSPLI GSYQGSQAPE RIESSGNSLF LAFRSDASVG
1251 LSGFAIEFKE KPREACFDPG NIMNGTRVGT DFKLGSTITY QCDSGYKILD
1301 PSSITCVIGA DGKPSWDQVL PSCNAPCGGQ YTGSEGVVLS PNYPHNYTAG
1351 QICLYSITVP KEFVVFGQFA YFQTALNDLA ELFDGTHAQA RLLSSLSGSH
1401 SGETLPLATS NQILLRFSAK SGASARGFHF VYQAVPRTSD TQCSSVPEPR
1451 YGRRIGSEFS AGSIVRFECN PGYLLQGSTA LHCQSVPNAL AQWNDTIPSC
1501 VVPCSGNFTQ RRGTILSPGY PEPYGNNLNC IWKIIVTEGS GIQIQVISFA
1551 TEQNWDSLEI HDGGDVTAPR LGSFSGTTVP ALLNSTSNQL YLHFQSDISV
1601 AAAGFHLEYK TVGLAACQEP ALPSNSIKIG DRYMVNDVLS FQCEPGYTLQ
1651 GRSHISCMPG TVRRWNYPSP LCIATCGGTL STLGGVILSP GFPGSYPNNL
1701 DCTWRISLPI GYGAHIQFLN FSTEANHDFL EIQNGPYHTS PMIGQFSGTD
1751 LPAALLSTTH ETLIHFYSDH SQNRQGFKLA YQGMEQQREP KPKSKYTSYM
1801 *
```

SEQ ID NO:11

```
5G-3V2 Protein sequence          2009 AA
   1  MEAIKTLSGI WNNINHVTSE EDTFIMYLGK PWLQVKIQVS QGGVALVSDM
  51  CPDPGIPENG RRAGSDFRVG ANVQFSCEDN YVLQGSKSIT CQRVTETLAA
 101  WSDHRPICRA RTCGSNLRGP SGVITSPNYP VQYEDNAHCV WVITTTDPDK
 151  VIKLAFEEFE LERGYDTLTV GDAGKVGDTR SVLYVLTGSS VPDLIVSMSN
 201  QMWLHLQSDD SIGSPGFKAV YQEIEKGGCG DPGIPAYGKR TGSSFLHGDT
 251  LTFECPAAFE LVGERVITCQ QNNQWSGNKP SCVFSCFFNF TASSGIILSP
 301  NYPEEYGNNM NCVWLIISEP GSRIHLIFND FDVEPQFDFL AVKDDGISDI
 351  TVLGTFSGNE VPSQLASSGH IVRLEFQSDH STTGRGFNIT YTTFGQNECH
 401  DPGIPINGRR FGDRFLLGSS VSFHCDDGFV KTQGSESITC ILQDGNVVWS
 451  STVPRCEAPC GGHLTASSGV ILPPGWPGYY KDSLHCEWII EAKPGHSIKI
 501  TFDRFQTEVN YDTLEVRDGP ASSSPLIGEY HGTQAPQFLI STGNFMYLLF
 551  TTDNSRSSIG FLIHYESVTL ESDSCLDPGI PVNGHRHGGD FGIRSTVTFS
 601  CDPGYTLSDD EPLVCERNHQ WNHALPSCDA LCGGYIQGKS GTVLSPGFPD
 651  FYPNSLNCTW TIEVSHGKGV QMIFHTFHLE SSHDYLLITE DGSFSEPVAR
 701  LTGSVLPHTI KAGLFGNFTA QLRFISDFSI SYEGFNITFS EYDLEPCDDP
 751  GVPAFSRRIG FHFGVGDSLT FSCFLGYRLE GATKLTCLGG GRRVWSAPLP
 801  RCVAECGASV KGNEGTLLSP NFPSNYDNNH ECIYKIETEA GKGIHLRTRS
 851  FQLFEGDTLK VYDGKDSSSR PLGTFTKNEL LGLILNSTSN HLWLEFNTNG
 901  SDTDQGFQLT YTSFDLVKCE DPGIPNYGYR IRDEGHFTDT VVLYSCNPGY
 951  AMHGSNTLTC LSGDRRVWDK PLPSCIAECG GQIHAATSGR ILSPGYPAPY
1001  DNNLHCTWII EADPGKTISL HFIVFDTEMA HDILKVWDGP VDSDILLKEW
1051  SGSALPEDIH STFNSLTLQF DSDFFISKSG FSIQFSTSIA ATCNDPGMPQ
1101  NGTRYGDSRE AGDTVTFQCD PGYQLQGQAK ITCVQLNNRF FWQPDPPTCI
1151  AACGGNLTGP AGVILSPNYP QPYPPGKECD WRVKVNPDFV IALIFKSFNM
1201  EPSYDFLHIY EGEDSNSPLI GSYQGSQAPE RIESSGNSLF LAFRSDASVG
1251  LSGFAIEFKE KPREACFDPG NIMNGTRVGT DFKLGSTITY QCDSGYKILD
1301  PSSITCVIGA DGKPSWDQVL PSCNAPCGGQ YTGSEGVVLS PNYPHNYTAG
1351  QICLYSITVP KEFVVFGQFA YFQTALNDLA ELFDGTHAQA RLLSSLSGSH
1401  SGETLPLATS NQILLRFSAK SGASARGFHF VYQAVPRTSD TQCSSVPEPR
1451  YGRRIGSEFS AGSIVRFECN PGYLLQGSTA LHCQSVPNAL AQWNDTIPSC
1501  VVPCSGNFTQ RRGTILSPGY PEPYGNNLNC IWKIIVTEGS GIQIQVISFA
1551  TEQNWDSLEI HDGGDVTAPR LGSFSGTTVP ALLNSTSNQL YLHFQSDISV
1601  AAAGFHLEYK TVGLAACQEP ALPSNSIKIG DRYMVNDVLS FQCEPGYTLQ
1651  GRSHISCMPG TVRRWNYPSP LCIATCGGTL STLGGVILSP GFPGSYPNNL
1701  DCTWRISLPI GYGAHIQFLN FSTEANHDFL EIQNGPYHTS PMIGQFSGTD
1751  LPAALLSTTH ETLIHFYSDH SQNRQGFKLA YQAYELQNCP DPPPFQNGYM
1801  INSDYSVGQS VSFECYPGYI LIGHPVLTCQ HGINRNWNYP FPRCDAPCGY
1851  NVTSQNGTIY SPGFPDEYPI LKDCIWLITV PPGHGVYINF TLLQTEAVND
1901  YIAVWDGPDQ NSPQLGVFSG NTALETAYSS TNQVLLKFHS DFSNGGFFVL
1951  NFHGQLIFTP LVKTENSMWC LLQCCPTPCF QLKFLDSAEG VYDSFALEAS
2001  VSCGPFFV*
```

SEQ ID NO:12

```
5G-3V3 Protein sequence              1784 AA
    1  MEAIKTLSGI WNNINHVTSE EDTFIMYLGK PWLQVKIQVS QGGVALVSDM
   51  CPDPGIPENG RRAGSDFRVG ANVQFSCEDN YVLQGSKSIT CQRVTETLAA
  101  WSDHRPICRA RTCGSNLRGP SGVITSPNYP VQYEDNAHCV WVITTTDPDK
  151  VIKLAFEEFE LERGYDTLTV GDAGKVGDTR SVLYVLTGSS VPDLIVSMSN
  201  QMWLHLQSDD SIGSPGFKAV YQEIEKGGCG DPGIPAYGKR TGSSFLHGDT
  251  LTFECPAAFE LVGERVITCQ QNNQWSGNKP SCVFSCFFNF TASSGIILSP
  301  NYPEEYGNNM NCVWLIISEP GSRIHLIFND FDVEPQFDFL AVKDDGISDI
  351  TVLGTFSGNE VPSQLASSGH IVRLEFQSDH STTGRGFNIT YTTFGQNECH
  401  DPGIPINGRR FGDRFLLGSS VSFHCDDGFV KTQGSESITC ILQDGNVVWS
  451  STVPRCEAPC GGHLTASSGV ILPPGWPGYY KDSLHCEWII EAKPGHSIKI
  501  TFDRFQTEVN YDTLEVRDGP ASSSPLIGEY HGTQAPQFLI STGNFMYLLF
  551  TTDNSRSSIG FLIHYESVTL ESDSCLDPGI PVNGHRHGGD FGIRSTVTFS
  601  CDPGYTLSDD EPLVCERNHQ WNHALPSCDA LCGGYIQGKS GTVLSPGFPD
  651  FYPNSLNCTW TIEVSHGKGV QMIFHTFHLE SSHDYLLITE DGSFSEPVAR
  701  LTGSVLPHTI KAGLFGNFTA QLRFISDFSI SYEGFNITFS EYDLEPCDDP
  751  GVPAFSRRIG FHFGVGDSLT FSCFLGYRLE GATKLTCLGG GRRVWSAPLP
  801  RCVAECGASV KGNEGTLLSP NFPSNYDNNH ECIYKIETEA GKGIHLRTRS
  851  FQLFEGDTLK VYDGKDSSSR PLGTFTKNEL LGLILNSTSN HLWLEFNTNG
  901  SDTDQGFQLT YTSFDLVKCE DPGIPNYGYR IRDEGHFTDT VVLYSCNPGY
  951  AMHGSNTLTC LSGDRRVWDK PLPSCIAECG GQIHAATSGR ILSPGYPAPY
 1001  DNNLHCTWII EADPGKTISL HFIVFDTEMA HDILKVWDGP VDSDILLKEW
 1051  SGSALPEDIH STFNSLTLQF DSDFFISKSG FSIQFSTSIA ATCNDPGMPQ
 1101  NGTRYGDSRE AGDTVTFQCD PGYQLQGQAK ITCVQLNNRF FWQPDPPTCI
 1151  AACGGNLTGP AGVILSPNYP QPYPPGKECD WRVKVNPDFV IALIFKSFNM
 1201  EPSYDFLHIY EGEDSNSPLI GSYQGSQAPE RIESSGNSLF LAFRSDASVG
 1251  LSGFAIEFKE KPREACFDPG NIMNGTRVGT DFKLGSTITY QCDSGYKILD
 1301  PSSITCVIGA DGKPSWDQVL PSCNAPCGGQ YTGSEGVVLS PNYPHNYTAG
 1351  QICLYSITVP KEFVVFGQFA YFQTALNDLA ELFDGTHAQA RLLSSLSGSH
 1401  SGETLPLATS NQILLRFSAK SGASARGFHF VYQAVPRTSD TQCSSVPEPR
 1451  YGRRIGSEFS AGSIVRFECN PGYLLQGSTA LHCQSVPNAL AQWNDTIPSC
 1501  VVPCSGNFTQ RRGTILSPGY PEPYGNNLNC IWKIIVTEGS GIQIQVISFA
 1551  TEQNWDSLEI HDGGDVTAPR LGSFSGTTVP ALLNSTSNQL YLHFQSDISV
 1601  AAAGFHLEYK TVGLAACQEP ALPSNSIKIG DRYMVNDVLS FQCEPGYTLQ
 1651  GRSHISCMPG TVRRWNYPSP LCIATCGGTL STLGGVILSP GFPGSYPNNL
 1701  DCTWRISLPI GYGAHIQFLN FSTEANHDFL EIQNGPYHTS PMIGQFSGTD
 1751  LPAALLSTTH ETLIHFYSDH SQNRQGFKLA YQA*
```

SEQ ID NO:13

```
5R-3V2 Protein sequence          2353 AA
   1 VGCAAGLGTG XSLRLALPSG DYLAPLIALP SSEEPSVSRA CGVSADMTAW
  51 RRFQSLLLLL GLLVLCARLL TAAKGQNCGG LVQGPNGTIE SPGFPHGYPN
 101 YANCTWIIIT GERNRIQLSF HTFALEEDFD ILSVYDGQPQ QGNLKVRLSG
 151 FQLPSSIVST GSILTLWFTT DFAVSAQGFK ALYEVLPSHT CGNPGEILKG
 201 VLHGTRFNIG DXIRYSCLPG YILEGHAILT CIVSPGNGAS WDFPAPFCRA
 251 EGACGGTLRG TSSSISSPHF PSEYENNADC TWTILAEPGD TIALVFTDFQ
 301 LEEGYDFLEI SGTEAPSIWL TGMNLPSPVI SSKNWLRLHF TSDSNHRRKG
 351 FNAQFQVKKA IELKSRGVKM LPSKDGSHKN SVLSQGGVAL VSDMCPDPGI
 401 PENGRRAGSD FRVGANVQFS CEDNYVLQGS KSITCQRVTE TLAAWSDHRP
 451 ICRARTCGSN LRGPSGVITS PNYPVQYEDN AHCVWVITTT DPDKVIKLAF
 501 EEFELERGYD TLTVGDAGKV GDTRSVLYVL TGSSVPDLIV SMSNQMWLHL
 551 QSDDSIGSPG FKAVYQEIEK GGCGDPGIPA YGKRTGSSFL HGDXLTFECP
 601 AAFELVGERV ITCQQNNQWS GNKPSCVFSC FFNFTASSGI ILSPNYPEEY
 651 GNNMNCVWLI ISEPGSRIHL IFNDFDVEPQ FDFLAVKDDG ISDITVLGTF
 701 SGNEVPSQLA SSGHIVRLEF QSDHSTTGRG XNITYTTFGQ NECHDPGIPI
 751 NGRRFGDRFL LGSSVSFHCD DGFVKTQGSE SITCILQDGN VVWSSTVPRC
 801 EAPCGGHLTA SSGVILPPGW PGYYKDSLHC EWIIEAKPGH SIKITFDRFQ
 851 TEVNYDTLEV RDGPASSSPL IGEYHGTQAP QFLISTGNFM YLLFTTDNSR
 901 SSIGFLIHYE SVTLESDSCL DPGIPVNGHR HGGDFGIRST VTFSCDPGYT
 951 LSDDEPLVCE RNHQWNHALP SCDALCGGYI QGKSGTVLSP GFPDFYPNSL
1001 NCTWTIEVSH GKGVQMIFHT FHLESSHDYL LITEDGSFSE PVARLTGSVL
1051 PHTIKAGLFG NFTAQLRFIS DFSISYEGFN ITFSEYDLEP CDDPGVPAFS
1101 RRIGFHFGVG DSLTFSCFLG YRLEGATKLT CLGGGRRVWS APLPRCVAEC
1151 GASVKGNEGT LLSPNFPSNY DNNHECIYKI ETEAGKGIHL RTRSFQLFEG
1201 DTLKVYDGKD SSSRPLGTFT KNELLGLILN STSNHLWLEF NTNGSDTDQG
1251 FQLTYTSFDL VKCEDPGIPN YGYRIRDEGH FTDTVVLYSC NPGYAMHGSN
1301 TLTCLSGDRR VWDKPLPSCI AECGGQIHAA TSGRILSPGY PAPYDNNLHC
1351 TWIIEADPGK TISLHFIVFD TEMAHDILKV WDGPVDSDIL LKEWSGSALP
1401 EDIHSTFNSL TLQFDSDFFI SKSGFSIQFS TSIAATCNDP GMPQNGTRYG
1451 DSREAGDTVT FQCDPGYQLQ GQAKITCVQL NNRFFWQPDP PTCIAACGGN
1501 LTGPAGVILS PNYPQPYPPG KECDWRVKVN PDFVIALIFK SFNMEPSYDF
1551 LHIYEGEDSN SPLIGSYQGS QAPERIESSG NSLFLAFRSD ASVGLSGFAI
1601 EFKEKPREAC FDPGNIMNGT RVGTDFKLGS TITYQCDSGY KILDPSSITC
1651 VIGADGKPSW DQVLPSCNAP CGGQYTGSEG VVLSPNYPHN YTAGQICLYS
1701 ITVPKEFVVF GQFAYFQTAL NDLAELFDGT HAQARLLSSL SGSHSGETLP
1751 LATSNQILLR FSAKSGASAR GFHFVYQAVP RTSDTQCSSV PEPRYGRRIG
1801 SEFSAGSIVR FECNPGYLLQ GSTALHCQSV PNALAQWNDT IPSCVVPCSG
1851 NFTQRRGTIL SPGYPEPYGN NLNCIWKIIV TEGSGIQIQV ISFATEQNWD
1901 SLEIHDGGDV TAPRLGSFSG TTVPALLNST SNQLYLHFQS DISVAAAGFH
1951 LEYKTVGLAA CQEPALPSNS IKIGDRYMVN DVLSFQCEPG YTLQGRSHIS
2001 CMPGTVRRWN YPSPLCIATC GGTLSTLGGV ILSPGFPGSY PNNLDCTWRI
2051 SLPIGYGAHI QFLNFSTEAN HDFLEIQNGP YHTSPMIGQF SGTDLPAALL
2101 STTHETLIHF YSDHSQNRQG FKLAYQAYEL QNCPDPPPFQ NGYMINSDYS
2151 VGQSVSFECY PGYILIGHPV LTCQHGINRN WNYPFPRCDA PCGYNVTSQN
2201 GTIYSPGFPD EYPILKDCIW LITVPPGHGV YINFTLLQTE AVNDYIAVWD
2251 GPDQNSPQLG VFSGNTALET AYSSTNQVLL KFHSDFSNGG FFVLNFHGQL
2301 IFTPLVKTEN SMWCLLQCCP TPCFQLKFLD SAEGVYDSFA LEASVSCGPF
2351 FV*
```

SEQ ID NO:14

PROTEIN SEQUENCE 5R23V2

```
LOCUS       5R23V2.PRO    2307 AA     PROT              UPDATED    05/11/101
DEFINITION  -
ACCESSION   -
KEYWORDS    -
SOURCE      -
FEATURES          From    To/Span      Description
    Peptide          1       2307      851 to 7771 of 5R23V2(translated)
ORIGIN      ?
        1 MTAWRRFQSL LLLLGLLVLC ARLLTAAKGQ NCGGLVQGPN GTIESPGFPH GYPNYANCTW
       61 IIITGERNRI QLSFHTFALE EDFDILSVYD GQPQQGNLKV RLSGFQLPSS IVSTGSILTL
      121 WFTTDFAVSA QGFKALYEVL PSHTCGNPGE ILKGVLHGTR FNIGDXIRYS CLPGYILEGH
      181 AILTCIVSPG NGASWDFPAP FCRAEGACGG TLRGTSSSIS SPHFPSEYEN NADCTWTILA
      241 EPGDTIALVF TDFQLEEGYD FLEISGTEAP SIWLTGMNLP SPVISSKNWL RLHFTSDSNH
      301 RRKGFNAQFQ VKKAIELKSR GVKMLPSKDG SHKNSVLSQG GVALVSDMCP DPGIPENGRR
      361 AGSDFRVGAN VQFSCEDNYV LQGSKSITCQ RVTETLAAWS DHRPICRART CGSNLRGPSG
      421 VITSPNYPVQ YEDNAHCVWV ITTTDPDKVI KLAXEEFELE RGYDTLTVGD AGKVGDTRSV
      481 LXVLTGSSVP DLIVSMSNQM WLHLQSDDSI GSPGFKAVYQ EIEKGGCGDP GIPAYGKRTG
      541 SSFLHGDXLT FECPAAFELV GERVITCQQN NQWSGNKPSC VFSCFFNFTA SSGIILSPNY
      601 PEEYGNNMNC VWLIISEPGS RIHLIFNDFD VEPQFDFLAV KDDGISDITV LGTFSGNEVP
      661 SQLASSGHIV RLEFQSDHST TGRGXNITYT TFGQNECHDP GIPINGRRFG DRFLLGSSVS
      721 FHCDDGFVKT QGSESITCIL QDGNVVWSST VPRCEAPCGG HLTASSGVIL PPGWPGYYKD
      781 SLHCEWIIEA KPGHSIKITF DRFQTEVNYD TLEVRDGPAS SSPLIGEYHG TQAPQFLIST
      841 GNFMYLLFTT DNSRSSIGFL IHYESVTLES DSCLDPGIPV NGHRHGGDFG IRSTVTFSCD
      901 PGYTLSDDEP LVCERNHQWN HALPSCDALC GGYIQGKSGT VLSPGFPDFY PNSLNCTWTI
      961 EVSHGKGVQM IFHTFHLESS HDYLLITEDG SFSEPVARLT GSVLPHTIKA GLXGNFTAQL
     1021 RFISDFSISY EGFNITFSEY DLEPCDDPGV PAFSRRIGFH FGVGDSLTFS CFLGYRLEGA
     1081 TKLTCLGGGR RVWSAPLPRC VAECGASVKG NEGTLLSPNF PSNYDNNHEC IYKIETEAGK
     1141 GIHLRTRSFQ LFEGDTLKVY DGKDSSSRPL GTFTKNELLG LILNSTSNHL WLEFNTNGSD
     1201 TDQGFQLTYT SFDLVKCEDP GIPNYGYRIR DEGHFTDTVV LYSCNPGYAM HGSNTLTCLS
     1261 GDRRVWDKPL PSCIAECGGQ IHAATSGRIL SPGYPAPYDN NLHCTWIIEA DPGKTISLHF
     1321 IVFDTEMAHD ILKVWDGPVD SDILLKEWSG SALPEDIHST FNSLTLQFDS DFFISKSGFS
     1381 IQFSTSIAAT CNDPGMPQNG TRYGDSREAG DTVTFQCDPG YQLQGQAKIT CVQLNNRFFW
     1441 QPDPPTCIAA CGGNLTGPAG VILSPNYPQP YPPGKECDWR VKVNPDFVIA LIFKSFNMEP
     1501 SYDFLHIYEG EDSNSPLIGS YQGSQAPERI ESSGNSLFLA FRSDASVGLS GFAIEFKEKP
     1561 REACFDPGNI MNGTRVGTDF KLGSTITYQC DSGYKILDPS SITCVIGADG KPSWDQVLPS
     1621 CNAPCGGQYT GSEGVVLSPN YPHNYTAGQI CLYSITVPKE FVVFGQFAYF QTALNDLAEL
     1681 FDGTHAQARL LSSLSGSHSG ETLPLATSNQ ILLRFSAKSG ASARGFHFVY QAVPRTSDTQ
     1741 CSSVPEPRYG RRIGSEFSAG SIVRFECNPG YLLQGSTALH CQSVPNALAQ WNDTIPSCVV
     1801 PCSGNFTQRR GTILSPGYPE PYGNNLNCIW KIIVTEGSGI QIQVISFATE QNWDSLEIHD
     1861 GGDVTAPRLG SFSGTTVPAL LNSTSNQLYL HFQSDISVAA AGFHLEYKTV GLAACQEPAL
     1921 PSNSIKIGDR YMVNDVLSFQ CEPGYTLQGR SHISCMPGTV RRWNYPSPLC IATCGGTLST
     1981 LGGVILSPGF PGSYPNNLDC TWRISLPIGY GAHIQFLNFS TEANHDFLEI QNGPYHTSPM
     2041 IGQFSGTDLP AALLSTTHET LIHFYSDHSQ NRQGFKLAYQ AYELQNCPDP PPFQNGYMIN
     2101 SDYSVGQSVS FECYPGYILI GHPVLTCQHG INRNWNYPFP RCDAPCGYNV TSQNGTIYSP
     2161 GFPDEYPILK DCIWLITVPP GHGVYINFTL LQTEAVNDYI AVWDGPDQNS PQLGVFSGNT
     2221 ALETAYSSTN QVLLKFHSDF SNGGFFVLNF HGQLIFTPLV KTENSMWCLL QCCPTPCFQL
     2281 KFLDSAEGVY DSFALEASVS CGPFFV*
```

SEQ ID NO:15

5R2_OC147 PROTEIN

```
LOCUS       TRANSLATIO    347 AA    PROT          UPDATED   05/11/101
DEFINITION  -
ACCESSION   -
KEYWORDS    -
SOURCE      -
FEATURES         From  To/Span     Description
  Peptide          1     347       851 to 1891 of 5r2_oc147 (translated)
ORIGIN      ?
        1 MTAWRRFQSL LLLLGLLVLC ARLLTAAKGQ NCGGLVQGPN GTIESPGFPH GYPNYANCTW
       61 IIITGERNRI QLSFHTFALE EDFDILSVYD GQPQQGNLKV RLSGFQLPSS IVSTGSILTL
      121 WFTTDFAVSA QGFKALYEVL PSHTCGNPGE ILKGVLHGTR FNIGDKIRYS CLPGYILEGH
      181 AILTCIVSPG NGASWDFPAP FCRAEGACGG TLRGTSSSIS SPHFPSEYEN NADCTWTILA
      241 EPGDTIALVF TDFQLEEGYD FLEISGTEAP SIWLTGMNLP SPVISSKNWL RLHFTSDSNH
      301 RRKGFNAQFQ VKKAIELKSR GVKMLPSKDG SHKNSVCESL SFLSED*
```

SEQ ID NO:16

5R2_AW PROTEIN

```
LOCUS       5R2_AW_PRO    372 AA    PROT           UPDATED   05/11/101
DEFINITION  -
ACCESSION   -
KEYWORDS    -
SOURCE      -
FEATURES          From  To/Span    Description
    Peptide         1      372     851 to 1966 of 5r2_aw (translated)
ORIGIN      ?
        1 MTAWRRFQSL LLLLGLLVLC ARLLTAAKGQ NCGGLVQGPN GTIESPGFPH GYPNYANCTW
       61 IIITGERNRI QLSFHTFALE EDFDILSVYD GQPQQGNLKV RLSGFQLPSS IVSTGSILTL
      121 WFTTDFAVSA QGFKALYEVL PSHTCGNPGE ILKGVLHGTR FNIGDKIRYS CLPGYILEGH
      181 AILTCIVSPG NGASWDFPAP FCRAEGACGG TLRGTSSSIS SPHFPSEYEN NADCTWTILA
      241 EPGDTIALVF TDFQLEEGYD FLEISGTEAP SIWLTGMNLP SPVISSKNWL RLHFTSDSNH
      301 RRKGFNAQFQ VKKAIELKSR GVKMLPSKDG SHKNSVWHQQ EFSKCRKKKR EIMTRNGRIS
      361 LTASGNLQFD N*
//
```

FIGURE 1

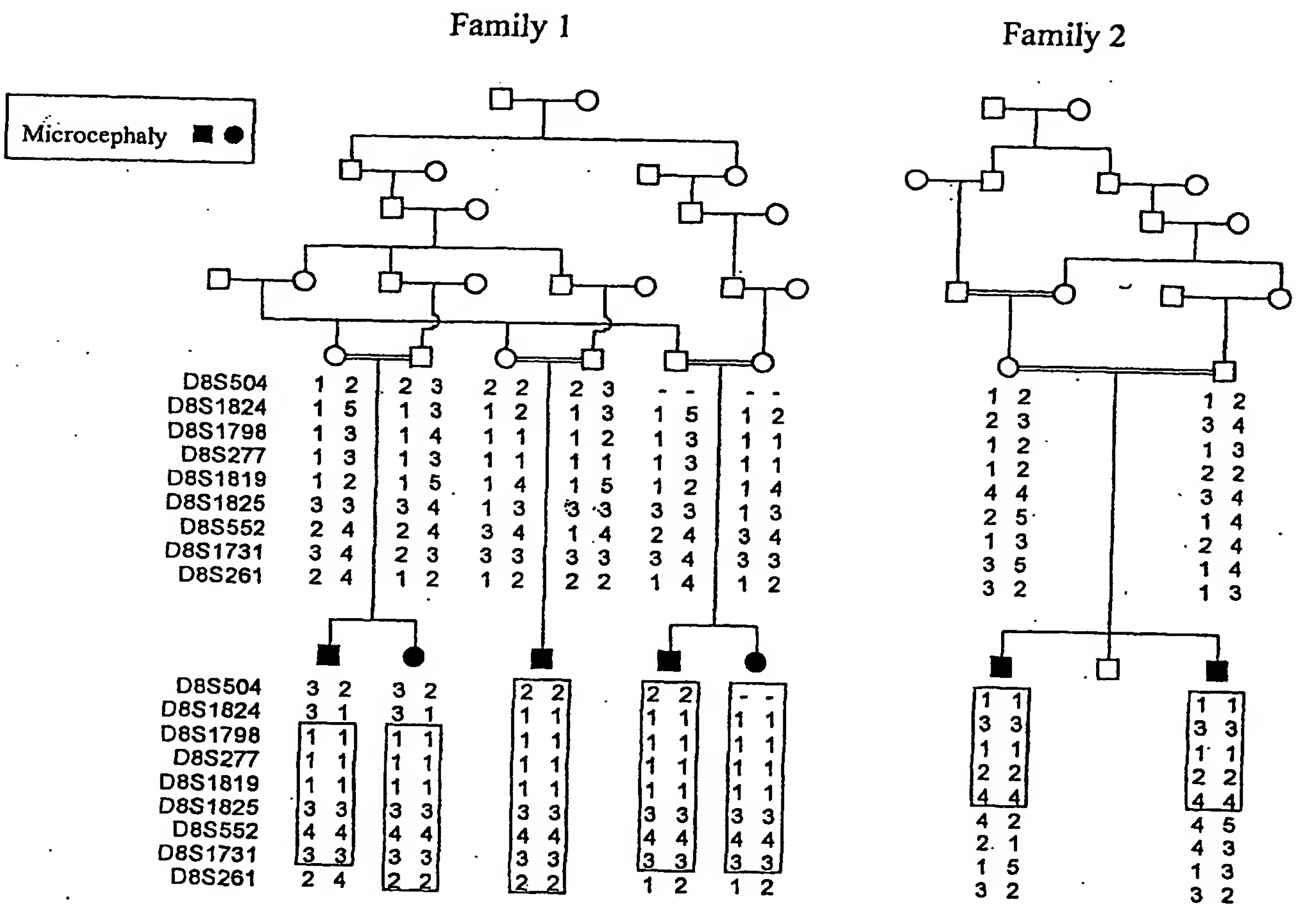

Family 1
Family 2
Microcephaly
D8S504
D8S1824
D8S1798
D8S277
D8S1819
D8S1825
D8S552
D8S1731
D8S261

FIGURE 2

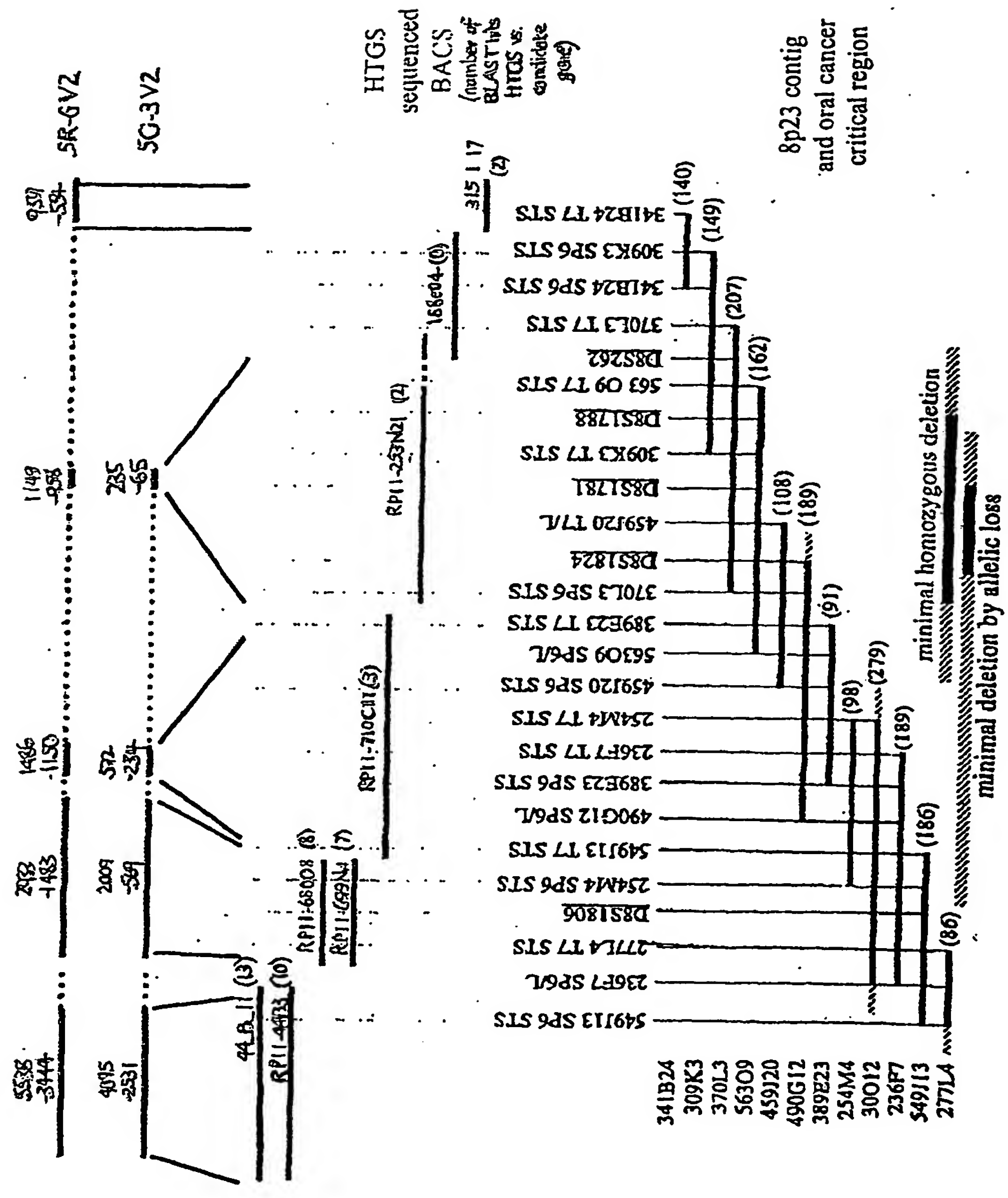
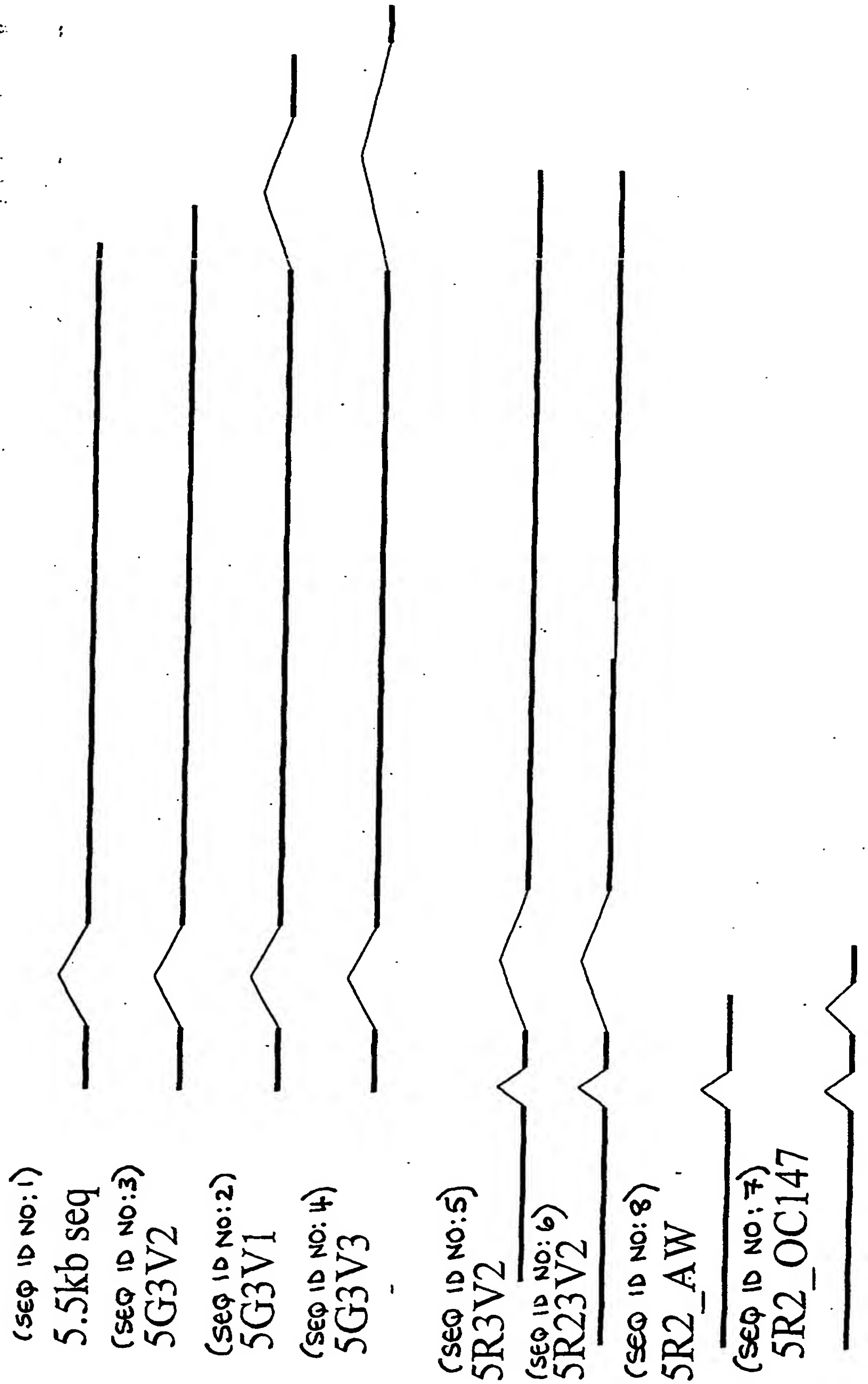

5R-6V2
5G-3V2
HTGS sequenced BACS (number of BLAST hits HTGS vs. candidate gene)
8p23 contig and oral cancer critical region
minimal homozygous deletion
minimal deletion by allelic loss
341B24 T7 STS
309K3 SP6 STS
341B24 SP6 STS
370L3 T7 STS
D8S262
563 O9 T7 STS
D8S1788
309K3 T7 STS
D8S1781
459J20 T7-L
D8S1824
370L3 SP6 STS
389E23 T7 STS
563O9 SP6-L
459J20 SP6 STS
254M4 T7 STS
236F7 T7 STS
389E23 SP6 STS
490G12 SP6-L
549J13 T7 STS
254M4 SP6 STS
D8S1806
277L4 T7 STS
236F7 SP6-L
549J13 SP6 STS
341B24
309K3
370L3
563O9
459J20
490G12
389E23
254M4
300O12
236F7
549J13
277L4

FIGURE 3


(SEQ ID NO:1)
5.5kb seq
(SEQ ID NO:3)
5G3V2
(SEQ ID NO:2)
5G3V1
(SEQ ID NO:4)
5G3V3
(SEQ ID NO:5)
5R3V2
(SEQ ID NO:6)
5R23V2
(SEQ ID NO:8)
5R2_AW
(SEQ ID NO:7)
5R2_OC147

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 01/02240

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C12N15/85 C12N15/86 C07K14/47 C07K16/18
C12Q1/68 G01N33/577 A61K31/713 A61K38/18 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K C12Q G01N A61K A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, PAJ, CAB Data, STRAND, BIOSIS, EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL SEQUENCE DATABASE 'Online! Hinxton, UK; 21 October 1999 (1999-10-21) NCI-CGAP: "xd71c12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2603062 3' mRNA sequence; EST" XP002175139 EMBL:AW104197, Comparison of Accession no. AW104197 and SEQ ID No. 8; abstract ---	1-6

-/--

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 August 2001	29/08/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hornig, H

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL SEQUENCE DATABSE 'Online! Hinxton, UK; 3 March 2000 (2000-03-03) L. HILLIER ET AL.: "zj96c05.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone ; EST" XP002175140 EMBL:AA705177, Comparison of Accession no. AA705177 and SEQ ID No.2; abstract	1-6
X	DATABASE EMBL SEQUENCE DATABASE 'Online! Hinxton, UK; 2 January 2000 (2000-01-02) K. KYUNG ET AL.: "Homo sapiens BAC clone RP11-221H10 from 8, complete sequence; HTG" XP002175141 EMBL:AC019176, Comparison of Accession no. AC019176 from position 13832-14773 and SEQ ID No. 6; abstract	1-6
A	SUN PAUL C ET AL: "Homozygous deletions define a region of 8p23.2 containing a putative tumor suppressor gene." GENOMICS, vol. 62, no. 2, 1 December 1999 (1999-12-01), pages 184-188, XP002175136 ISSN: 0888-7543 cited in the application the whole document	
A	ISHWAD CHANDRAMOHAN S ET AL: "Frequent allelic loss and homozygous deletion in chromosome band 8p23 in oral cancer." INTERNATIONAL JOURNAL OF CANCER, vol. 80, no. 1, 5 January 1999 (1999-01-05), pages 25-31, XP002175137 ISSN: 0020-7136 the whole document	
A	SUNWOO JOHN B ET AL: "Localization of a putative tumor suppressor gene in the sub-telomeric region of chromosome 8p." ONCOGENE, vol. 18, no. 16, 22 April 1999 (1999-04-22), pages 2651-2655, XP001015856 ISSN: 0950-9232 the whole document	

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
T	SUN PAUL C ET AL: "Transcript map of the 8p23 putative tumor suppressor region." GENOMICS, vol. 75, no. 1-3, July 2001 (2001-07), pages 17-25, XP002175138 ISSN: 0888-7543 the whole document -----	

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claim 20 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
Although claim(s) 18 and 19 (as far as in vivo methodds are concerned) are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.